DOCUMENT RESUME

ED 219 236 SE 038 287

AUTHOR	Jensen, R. S.
TITLE	First Symposium on Aviation Psychology. Proceedings of the Symposium on Aviation Psychology (Columbus, Ohio, April 21 and 22, 1981).
INSTITUTION	Ohio State Univ., Columbus. Aviation Psychology Lab.
SPONS AGENCY	Battelle Memorial Inst., Columbus, Ohio. Columbus Labs.; National Aeronautics and Space Administration, Mountain View, Calif. Ames Research Center.
REPORT NO	APL-TR-1-81
PUB DATE	Apr 81
NOTE	419p.; Also sponsored by the Association of Aviation Psychologists.
EDRS PRICE	MF01/PC17 Plus Postage.
DESCRIPTORS	*Aerospace Education; *Aircraft Pilots; Aviation Technology; *Flight Training; Higher Education; *Human Factors Engineering; *Performance Factors; *Psychology; Visual Perception
IDENTIFIERS	*Aviation Psychology

ABSTRACT

This volume contains five complete manuscripts and two abstracts presented, and three papers submitted but not presented, at this symposium on aviation psychology. The objective of the symposium was to critically examine the impact of high technology on the role, responsibility, authority, and performance of human operators in modern aircraft and air traffic control systems. Papers are grouped by the following subject area categories: 1) cockpit monitoring concepts; 2) cockpit information systems; 3) pilot judgment; 4) vision and visual perception; 5) crew workload, coordination, and complement; 6) pilot selection; 7) pilot training; and 8) performance assessment. (Author/JN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219236

TECHNICAL REPORT: APL-1-81

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

First Symposium on Aviation Psychology

April 21 and 22, 1981

The Aviation Psychology Laboratory
The Ohio State University
Columbus, Ohio

Convener: R. S. Jensen

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Aviation Psychology
Laboratory
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Sponsored by:

The NASA Ames Research Center
The Association of Aviation Psychologists
Battelle, Columbus Laboratories

SE 038 287

Proceedings of the

SYMPOSIUM ON AVIATION PSYCHOLOGY

April 21 and 22, 1981

The Aviation Psychology Laboratory
The Ohio State University
Columbus, Ohio

Convener: R. S. Jensen

Sponsored by:

The NASA Ames Research Center
The Association of Aviation Psychologists
Battelle, Columbus Laboratories

FORWARD

This volume contains the proceedings of the First Symposium on Aviation Psychology conducted by the Aviation Psychology Laboratory of the Ohio State University in Columbus Ohio on April 21 and 22, 1981 sponsored by the NASA Ames Research Center, The Association of Aviation Psychologists, and Battelle, Columbus Laboratories. The Technical Monitor was Dr. John Lauber of the NASA Ames Research Center. It contains complete manuscripts of most of the papers presented at the meeting and abstracts of the others. It also contains three papers submitted for the proceedings and not presented at the meeting. The papers were grouped by subject areas closely following their order of presentation in the program.

The objective of this symposium was to critically examine the impact of high technology on the role, responsibility, authority, and performance of human operators in modern aircraft and air traffic control systems. Our theme was "Aviation Psychology since Paul Fitts, Is Advancing Technology Ignoring Human Performance in Aviation Systems?" Human engineering principles set forth by Paul Fitts for aviation systems were used as the basis for an examination of modern ground and airborne display and control concepts as they relate to human perceptual, motor, and decisional performance, operator selection and training requirements, and crew coordination.

The role of the human operator in man-machine systems has been changing throughout the history of automation. Because new systems frequently require information processing rates and prediction accuracies far exceeding man's capabilities, a tempting alternative is to limit man's role to supervisor and to use a servo as the active control element. Generally, it is more difficult to find solutions that enhance man's capabilities as the system controller. Furthermore, because of their lack of experience with human information processing systems, engineers are less inclined to seek such solutions (Singleton, 1976). Consequently, man is being given a supervisory role consisting of planning, teaching, monitoring, and intervening (Sheridan, 1976).

One of the best examples of the changing role of the human operator in a man-machine system is that of the pilot of a modern airplane. Continuing demands for improved safety, efficiency, energy conservation, and noise reductions with increasing traffic flow have led to increasingly complex systems and control tasks. More and more functions are being handled automatically by ground-based and airborne computing systems, and the pilot is taking the role of a system supervisor who exercises "control by exception" authority only. Nevertheless, despite this increasing role of automation, the pilot remains a redundant system element responsible for manual takeover in the "exceptional" event of partial system failure or other unpredictable contingency that requires improvisation.

In actual practice, the pilot's role as a redundant system element is extremely important. The autopilot is useful during the many "hours of boredom," relieving the pilot of needless attention to aircraft control tasks. However, the autopilot has not been very useful during the "moments of stark terror"(Kennelly, 1970). At the first indication of unusual circumstances (e.g., traffic avoidance, frequent flight path changes, partial system failure, turbulence penetration, passenger discomfort, wind shear, etc.) the pilot's initial action is to disengage the autopilot, whether or not such action is needed. Thus, the autopilot has proved to be most used when the pilot workload levels are low and least used during many periods of high cockpit workload.

In a 1951 report for the NRC entitled, "Human Engineering for an Effective Air-Navigation and Air Traffic-Controller System," Paul Fitts set forth a number of longstanding principles concerning the effective allocation of tasks to men and machines that are studied in human factors classrooms to this day. Among the principles established by Fitts and his colleagues were the following:

1. Human tasks should provide activity.
2. Human tasks should be intrinsically motivating.
3. Machines should monitor humans, not the converse.

Although the tasks of pilots and air traffic controllers at that time were largely "manual" in comparison to today, Fitts could foresee the possibility of conflicts in man-machine task allocations as automation developed.

In our day, the unquestioned motivation behind virtually every technological advancement in the cockpit is "workload reduction". As a result, we have combination control-wheel steering, auto-throttle, and autopilot systems that permit the pilot to assume control of the system at any level in the control hierarchy. A pilot can program his flight on the runway in Paris, take off and touch only push-button controls until he taxis off the runway in New York. His "workload" is "reduced" under normal flying conditions to the level of a living room observer of Monday night football.

As a result of these "advances", the task assigned to the pilot may be inadequate considering the Fitts principles. The pilot's task requires almost no physical activity, it fails to be intrinsically motivating, and it amounts to a task of monitoring a machine rather than the converse. Thus, the only conditions under which the pilot is overloaded are those cases in which his equipment is degraded. The effect may have been to reduce the pilot's task in normal conditions to a level beneath what Fitts considered adequate without helping and perhaps even hurting his manual control capabilities during flight under degraded conditions.

In addition to the problems of continuous control that are introduced, automation tends to change the requirements for complex decision-making, operator selection and training, and crew coordination. There is a real need at this time for a critical examination of the impact on our aviation system of "engineering solutions" before they find a "problem" that may not exist. The 1981 Symposium on Aviation Psychology initiated this examination in a series of paper sessions given by experts in the field.

Richard S. Jensen

CONTENTS

FORWARD i

PROGRAM viii

INTRODUCTION

The Role of Communications, Socio-Psychological, and Personality Factors in the Maintenance of Crew Coordination 1
H.C. Foushee

Monitors of Human Performance 12
J.W. Senders

Human Factors and Aviation Safety: A Program of Research on Human Factors in Aviation 21*
S.N. Roscoe

COCKPIT MONITORING CONCEPTS

Fitts' Principles Still Applicable: Computer Monitoring of Fighter Aircraft Emergencies 28
J. Reising
L. Hitchcock

The Performance of Warning Systems in Avoiding Controlled Flight Into Terrain (CFIT) Accidents 38
J.P. Loomis
R.F. Porter

Assessing Emergency Interface Design 51
W.C. Allen

The Effects of Alert Priortization and Inhibit Logic on Pilot Performance 61**
D.A. Po-Chedley

PROCRU: A Model for Analyzing Flight Crew Procedures in Approach to Landing 62**
S. Baron
G. Zacharias
R. Muralidharan
R. Lancroft

* Not Presented at Meeting
** Abstract Only

COCKPIT INFORMATION SYSTEMS

Ergonomic Aspects of Cockpit Lay-Out................................63
H. Aubauer
W. Sperr

A Comparison of Tracking with
Visual and Kinesthetic-Tactual Displays.............................74
R.J. Jagacinski
J.M. Flach
R.D. Gilson

General Aviation Cockpit Design Features
Related to Inadvertant Landing Gear Retraction Accidents............84
A. Diehl

In-Trail Following During Profile Descents
with a Cockpit Display of Traffic Information.......................94
S.L. Chappell
E.A. Palmer

Pave Low III: Interior Lighting Reconfiguration
for Night Lighting and Night Vision Goggle Compatability...........106
H.L. Task
L.L. Griffin

Preliminary Evaluation of an
On-Board Computer-Based Information System........................117**
S.H. Rouse
W.B. Rouse

Intelligibility of and Pilot's Reactions
to Various Types of Synthesized Speech............................118**
G. Rothbauer (Presented by E. Gora)

Head-Up Displays in Operation:
Some Unanswered Questions...119**
R.L. Newman
W.L. Welde

PILOT JUDGMENT

A Study of Decision-Making Behavior of
Aircraft Pilots Deviating from a Planned Flight....................120
G.W. Flathers, II
W.C. Giffin
T.H. Rockwell

An Analysis of Aircrew Procedural Compliance......................134**
J.E. Schofield
W.C. Giffin

7

Instructional Design for Aircrew Judgment Training................145
F. Brecke

Airmanship: An Introduction......................................161
A. Mane

Developing a Civil Aviation Pilot
Judgment Training and Evaluation Manual..........................166
J. Berlin
C. Holmes

Planning Behavior of Pilots
in Abnormal and Emergency Situations..............................171**
G. Johannsen
W.B. Rouse

VISION AND VISUAL PERCEPTION

Landing Airplanes, Detecting Traffic, and the Dark Focus..........172
S.N. Roscoe (Presented by R. Hennessy)

The Dark Focus of Accomodation
and Pilot Performance...182
R.A. Benel
T.L. Amerson, Jr.

Functional Optical Invariants:
A New Methodology for Aviation Research...........................192
R. Warren
D.H. Owen

Fractional Rates of Change as Functional Optical Invariants........205
S.J. Mangold
D.H. Owen
R. Warren

CREW WORKLOAD, COORDINATION, AND COMPLEMENT

Visual Scanning Behavior and Pilot Performance....................216
R.L. Harris, Jr.
J.R. Tole
A.T. Stephens
A.R. Ephrath

Tanker Avionics and Aircrew Complement Evaluation.................226
R.W. Moss
G.J. Barbato

Operational Monitoring in Multi-Crew Transport Operations..........247**
H.W. Orlady

8

An Organization Development Approach
to Resource Management in the Cockpit..............................248
L.O. Rings

PILOT SELECTION

Validation of a Proposed Pilot Trainee Selection System............255
J.M. Koonce

Sex as a Moderator Variable in the Selection
and Training of Persons for a Skilled Task.........................261
T.M. McCloy
J.M. Koonce

Changes in the U.S. Army Aviator Selection
and Training Program...267
W.R. Brown
J.A. Dohme
M.G. Sanders

Individual Differences in Multi-Task Response Strategies...........279
D. Damos
T. Smist

PILOT TRAINING

Simulation Technology and the Fixation Phase.......................289
E.A. Stark

An Adaptive Private Pilot Pilot Certification Exam.................306
S.R. Trollip
R.I. Anderson

Towards an Internal Model in Pilot Training........................316
R.J. Braune
S.R. Trollip

The Tomorrow Learning Machine......................................326
W.W. Castor

Operator Skill Retention in Automated Systems......................334**
D.B. Beringer

PERFORMANCE ASSESSMENT

Measuring Pilot Air Combat Maneuvering Performance.................335*
E.E. Eddowes

Performance Measurement and the Navy's
Tactical Aircrew Training System (TACTS)...........................343
G.R. Stoffer

Development and Application of Air Combat Performance Assessment Methods....................................362
A.P. Ciavarelli
A.M. Williams
C.A. Brictson

Measures of Effectiveness in Evaluating a Prototype General Aviation In-Flight Simulator..............................376
B. Strauch

LIST OF ATTENDEES..391

AUTHOR INDEX...401

1981 SYMPOSIUM ON AVIATION PSYCHOLOGY

at the Holiday Inn "on the Lane"
Columbus, Ohio

Sponsored by

The OSU Aviation Psychology Laboratory
The Association of Aviation Psychologists
The NASA Ames Research Center

Monday, April 20

19:00	Reception -- Clark Room Sponsored by Battelle, Columbus Laboratories	

Tuesday, April 21

Plenary Session
Moderator: Dr. Richard Jensen, Director
The Aviation Psychology Laboratory
Room: Sheridan and Custer

08:45	Opening Remarks	Dr. Marion Smith, Associate Dean, College of Engineering Dr. Ann Reynolds, Provost
09:15	Keynote Address "Aviation Psychology Since Paul Fitts"	Dr. Conrad Kraft Boeing Aerospace Company
10:00	Coffee	
10:30	Invited Address "Monitors of Human Performance"	Dr. John Senders University of Toronto
11:00	Invited Address "Within Cockpit Communication Patterns and Flight Crew Performance"	Dr. Clayton Foushee NASA Ames Research Center
12:30	Lunch	

* *

Training I
Moderator: Dr. Merv Strickler, Former Director of the FAA Aviation Education Programs Division
Room: Sheridan

13:00	"Simulation Technology and the Fixation Stage"	Dr. Ed Stark Singer/Link
13:30	"Development and Application of Air Combat Performance Assessment Methods"	Mr. Anthony Ciavarelli Dunlap & Associates

Tuesday, April 21, Con't.

13:50	"The Navy's Tactical Air Combat Training System (TACTS)"	Lt. Gerald Stoffer Naval Training and Equipment Center
14:10	"The Air Force's Simulator for Air-to-Air Combat (SAAC)"	Lt. Col. Joe Robinson Luke Air Force Base
14:30	"Operator Skill Retention in Automated Systems"	Dr. Dennis Beringer University of Wisconsin
15:00	Coffee	

* *

Cockpit Information Systems: Models, Displays, and Controls
Moderator: Dr. John Riesing, AF Flight Dynamics Laboratory
Room: Custer

13:00	"1951 - 1981: A Personal Perspective"	Dr. Malcolm Ritchie Wright State University
13:30	"PROCRU: A Model for Analyzing Flight Crew Procedures in Approach to Landing"	Dr. Sheldon Baron Bolt, Beranek and Newman
13:50	"Intrail Following During Profile Descents with a Cockpit Display of Traffic Information"	Ms. Sherry Chappell Dr. Everett Palmer NASA Ames Research Center
14:10	"Preliminary Evaluation of an On-board Computer-based Information System"	Ms. Sandra Rouse Dr. Bill Rouse University of Illinois
14:30	"General Aviation Cockpit Design Features Related to Inadvertent Landing-Gear Retraction Accidents"	Dr. Al Diehl FAA, Washington Office
15:00	Coffee	

* *

Vision - Visual Perception
Moderator: Dr. Dean Owen, Ohio State University
Room: Sheridan

15:30	"Landing Airplanes, Detecting Traffic, and the Dark Focus"	Dr. Stan Roscoe New Mexico State Univ.
15:50	"The Dark Focus of Accommodation and Pilot Performance"	Dr. Russ Benel Dr. Thomas Amerson, Jr. The Essex Corp.

Tuesday, April 21, Con't.

16:10	"Functional Optical Invariants: A New Methodology for Aviation Research"	Dr. Rik Warren Dr. Dean Owen Ohio State University
16:30	"Fractional Rates of Change as Functional Optical Invariants"	Ms. Sue Mangold Dr. Dean Owen Dr. Rik Warren Ohio State University
16:50	"ATC System-Induced Pilot Error: Human Factors and Legal Considerations"	Mr. Frank Fowler Fowler and Associates

* *

Pilot Judgment I
Moderator: Dr. Jerry Berlin, Embry Riddle Aeronautical University
Room: Custer

15:30	"Instructional Design for Aircrew Judgment Training"	Dr. Fritz Brecke Veda, Inc.
15:50	"Planning Behavior of Pilots in Abnormal and Emergency Situations"	Dr. Gunnar Johannsen Institut fur Anthropotechnik, Germany Dr. Bill Rouse University of Illinois
16:10	"Decision Making During Critical Inflight Events"	Mr. Bill Flathers, MITRE Dr. Tom Rockwell Dr. Walt Giffin Ohio State University
16:30	"Airmanship - An Introduction"	Mr. Amir Mane University of Illinois
16:50	"A Civil Aviation Training Program to Improve Pilot Judgment"	Dr. Jerry Berlin Dr. Charles Holmes Embry Riddle Aeronautical University

* *

19:00 Banquet at Holiday Inn in Sherman Room
Speaker: Dr. Richard Anderson
Professor of Geology and Mineralogy
The Ohio State University
Title: "Energy Alternatives"

Wednesday, April 22

Cockpit Monitoring Concepts
Moderator: Dr. Malcolm Ritchie
Room: Sheridan

08:30	"Fitt's Principles Still Applicable: Computer Monitoring of Cockpit Emergencies"	Dr. John Reising W-P Air Force Base
08:50	"Intelligibility of and Pilots' Reactions to Various Types of Synthesized Speach"	Dr. Evelyn Gora Technical University of Munich
09:10	"A Retrospective Examination of the Performance of Warning Devices in Avoiding Controlled-Flight-Into Terrain (CFIT) Accidents"	Mr. Jim Loomis and Mr. R. F. Porter Battelle, Columbus Laboratories
09:30	"The Effects of Alert Prioritization and Inhibit Logic on Pilot Performance"	Mr. David Po-Chedley Douglas Aircraft Corp.
09:50	"Computer Aided Decision Making"	Mr. Bill Allen Stanford University
10:10	Coffee	
10:30	"A Comparison of Tracking with Visual and Kinesthetic Tactual Displays"	Dr. Richard Jagacinski Mr. John Flach, Dr. Richard Gilson Ohio State University
10:50	"Ergonomics Aspects in Cockpit-Layout"	Dr. Walter Sperr, Consultant Mech. Engineering Vienna Dr. Helfried Aubauer Hinterbruhl, Austria
11:10	"PAVE LOW III Interior Lighting Reconfiguration for Night Vision Goggle Compatibility"	Dr. Lee Griffin W-P Air Force Base
11:30	"Head Up Displays in Operation: Some Unanswered Questions"	Mr. Richard Newman Crew Systems Consultants Mr. Bill Welde AFAMRL
11:50	"Uses of Stereographic Displays in Aircraft Cockpits"	Ms. S. Joy Mountford Mr. Ben Somberg Honeywell

Wednesday, April 22, Con't.

Training II

Moderator: Dr. Stanley N. Roscoe, New Mexico State University
Room: Custer

08:30	"Adaptive Models in Training"	Dr. Stan Trollip Mr. Richard Anderson University of Illinois
09:00	"Towards an Internal Model in Pilot Training"	Mr. B. Braune Dr. Stan Trollip University of Illinois
09:30	"The Tomorrow Learning Machine"	Mr. Webb Caster Aviation Simulation Technology, Inc.
10:00	Coffee	
10:30	"Measures of Effectiveness to Evaluate a Prototype GA Inflight Simulator"	Dr. Berry Strauch Embry Riddle Aeronautical University
10:50	"Computer Modeling of Realistic Terrain Models"	Dr. Chuck Csuri Ohio State University

* *

Pilot Judgment II

Moderator: Dr. Jerry Berlin, Embry Riddle Aeronautical University
Room: Sherman

10:30	Round Table Discussion "Pilot Judgment Training and Evaluation"	W. Flathers, MITRE R. Jensen, OSU T. Rockwell, OSU W. Giffin, OSU F. Brecke, Veda, Inc. R. Benel, Essex Corp. M. Strickler A. Diehl, FAA

* *

Pilot Selection

Moderator: Dr. Sergie Kochkin, United Airlines
Room:

13:30	"Individual Differences in Multi-Task Response Strategies"	Dr. Diane Damos University of Oregon

13:50	"Validation of a Proposed Pilot Trainee Selection System"	Lt. Col. Jeff Koonce Air Force Academy
14:10	"Sex as a Moderator Variable in the Selection and Training of Persons for Learning Flight Maneuvers"	Maj. Tom McCloy The Air Force Academy
14:30	"Changes in the U.S. Army Aviation Selection and Training Program	Mr. William Brown Dr. J. A. Dohme Dr. M. G. Sanders U.S. Army Research Institute

* *

Crew Workload, Coordination, and Complement
Moderator: Dr. Clayton Foushee, NASA Ames Research Center
Room:

13:30	"Mental Workload and Visual Scanning"	Dr. Randel Harris NASA Langley Dr. J.R. Tole Dr. A.T. Stephens Dr. A.E. Ephoth MIT
13:50	"Tanker Avionics/Aircrew Complement Evaluation"	Dr. Richard Moss W-P Air Force Base
14:10	"Operational Monitoring in Multi-Crew Transport Operation"	Capt. Harry Orlady United Airlines (Ret.)
14:30	"An Organization Development Approach to Resource Management in the Cockpit"	Mrs. Linda Orlady Rings Ohio State University
15:00	Open House	

THE AVIATION PSYCHOLOGY LABORATORY
355 Baker Systems Engineering

Dave Park
Karl Olson
Dave Smith
Larry Hettinger
Diane Rush
Greg Alexander

THE ROLE OF COMMUNICATIONS, SOCIO-PSYCHOLOGICAL, AND PERSONALITY FACTORS IN THE MAINTENANCE OF CREW COORDINATION

H. Clayton Foushee, Ph.D.
National Research Council
National Aeronautics and Space Administration
Ames Research Center
Moffett Field, California 94035

There is increasing evidence that many air transport incidents and accidents are the result of the improper utilization of the resources accessible to flight deck crew members. These resources obviously include the hardware and technical information necessary for the safe and efficent conduct of the flight, but they also include the human resources which must be coordinated effectively. Today, I want to focus on these human resources, and try to describe how communications styles, socio- psychological factors, and personality characteristics can affect crew coordination.

A number of air carriers are implementing resource management training programs in an attempt to create an awareness among pilots that the proper coordination and utilization of human resources are essential to continued safe operations in increasingly complex environments. With this realization has come a growing sense that technical proficiency alone is not enough.

It is clear that a consideration of flight crews as small groups suggests that a number of social and personality factors are relevant to crew performance. Such factors as leadership, group relations, and communications patterns have been consistently linked to performance in a variety of small group performance settings. There is no reason to believe that the flight deck is immune to the influence of group dynamics.

Often, I am going to refer to the interaction of social and personality characteristics so if I may, I would like to describe exactly why we feel that this interaction is so important. Personality psychology has been criticized because of its inability to demonstrate convincingly strong relationships with various types of behavior. Social psychology has suffered analogous difficulties in documenting consistent relationships between social factors and behavior. When you talk to a social or personality psychologist about the dynamics of a given relationship, somewhere in the discussion I can almost guarantee that he or she is going to say, "Well, it all depends." Part of this problem stems from the fact that a personality profile associated with good performance in one situation may be related to marginal performance in another. In the same way, certain social factors may affect one personality type one way and another entirely differently. Thus, we are faced with the necessity of looking at the joint contribution of social and personality factors to a certain outcome.

Anyone vaguely familiar with psychology probably knows the plethora of tendencies cited as relatively stable personality characteristics. Many have been explored, however, two dimensions seem particularly relevant to the small group situation and have been extensively validated in many situations by various researchers (e.g. Spence & Helmreich, 1978). The first is usually referred to as instrumentality or goal orientation. Persons high on this dimension tend to be highly performance oriented, decisive, capable of getting the job done, and so forth. The second dimension is often called expressivity or group orientation. Highly expressive people tend to be sensitive to the feelings of others, warm in interpersonal relationships, and communicative with people. Another characteristic of these two dimensions is that they are relatively orthogonal --that is, high on one does not necessarily imply low on the other and, of course, vice-versa.

With respect to the flight deck, one would obviously expect a strong positive relationship between goal orientation and performance, while predicting a rather meager or non-existent relationship between group orientation and performance. However, it should be apparent that those high in group orientation would establish warmer, more pleasant working relationships with other flight deck crew members.

Several recent accidents suggest that under some circumstances, the lack of an effective group orientation and inadequate communications, coupled with the obvious role structure of the flight deck, are contributory factors. Subordinate crew members have often complained to me that captains can sometimes be so insensitive and intimidating that they are hesitant to speak up even in potentially dangerous situations. Lest anyone have any doubts about the existence of overbearing captains, I would like to read the following anonymous report submitted to the NASA/FAA Aviation Safety Reporting System by a first officer who did speak up:

> I was the first officer on an airlines flight into Chicago O'Hare. The captain was flying, we were on approach to 4R getting radar vectors and moving along at 250 knots. On our approach Approach Control told us to slow to 180 knots. I acknowledged and waited for the captain to slow down. He did nothing, so I figured he didn't hear the clearance. So I repeated, "Approach" said slow to 180," and his reply was something to the effect of, "I'll do what I want." I told him at least twice more and received the same kind of answer. Approach Control asked us why we had not slowed yet. I told them we were doing the best job we could and their reply was, "You almost hit another aircraft." They then asked us to turn east. I told them we would rather not because of the weather and we were given present heading and to maintain 3000 ft. The captain descended to 3000 ft. and kept going to 2500 ft. even though I told him our altitude was 3000 ft. His comment was, "You just look out the damn window."

There is a danger inherent in such situations that first officers can become "conditioned" into not speaking up after running into a number of captains such as the one in the last report. This can extend even to relationships with captains who do encourage open channels of communication, as in the following:

> I was the copilot on a flight from JFK to BOS. The captain was flying. Departure turned us over to center and we were given FL210 which was our flight plan altitude. I noted that we had reached FL210 and were continuing through it, but was reluctant to say anything. As we climbed through 21,300 ft., I mentioned it to the captain, but not forcefully enough, and he did not hear me. I mentioned it again and pointed to the altimeter. We were at 21,600 ft. when the climb was stopped, and we descended back to 21,000. As we started our descent, center called and told us to maintain FL210. The captain said he had misread his altimeter and thought he was 1000 ft. lower than he was. I believe the main factor involved here was my reluctance to correct the captain. This captain is very "approachable" and I had no real reason to hold back. It is just a bad habit that I think a lot of copilots have of double-checking everything before we say anything to the captain.

It should come as no great surprise to anyone that this situation can produce, and has produced, disastrous consequences. In a 1979 crash of a north-eastern commuter carrier, the first officer failed to take over control of the aircraft when the captain apparently became incapacitated. The captain was a company vice- president, the first officer, a newly hired pilot on probation. The captain, according to reports, was a gruff personality and was observed to be visibly upset on the day of the accident. Further, this captain apparently had a history of not acknowledging callouts. Now I suspect that about the last thing that that particular first officer would have been willing to do with that particular captain was take over control of his aircraft. It would appear that the first officer was intimidated and that the accident might not have occurred given a different orientation by the captain.

In another, a twin-jet slid off the end of the runway after crossing the outer marker approximately 60 knots too fast. Although the captain was apparently not aware of the excessive speed, evidence indicates that the first officer knew, but could only muster a rather sheepish comment about the possible presence of a tail-wind. I obviously cannot state that the group orientations of these captains were causal, but we can reasonably speculate that warmer, more sensitive interpersonal styles on the part of captains, in general, may increase the probability that subordinate crew members will speak up. Apparently, this reluctance to question captains or assume control is not just an isolated problem. In a study conducted by Dr. C. R. Harper at United Air Lines, captains feigned subtle incapacitation at a predetermined point during final approach in simulator trials. In that study, roughly one-third of the aircraft "hit the ground" because for some reason, the first officers did not take control. This finding is disturbing despite the fact that "simulator complacency" may account for some of these occurences.

These observations also raise questions about within-cockpit communications patterns in general. There is a feeling among human factors specialists, airline training departments, and social and personality psychologists that communications patterns exert significant influences on important performance-related factors. It is also fairly clear that communications styles are rooted in the realm of personality. At the very least, communications patterns are crucial determinants of information transfer, but research has shown that they are also related to such factors as group cohesion, attitudes toward work, and complacency. That communications difficulties arise in the cockpit is not surprising in light of the fact that some large carriers employ thousands of pilots, who in many cases have never met prior to a trip. Thus, responsibilities that may be implicitly understood in crews who have flown together frequently, have to be explicitly delineated in those that have not.

I was provided the opportunity for a more systematic look at cockpit communications patterns by the data acquired in a NASA full-mission simulation study conducted by the late Pat Ruffell Smith (1979) and NASA researchers. In that study, fully qualified B-747 crews flew a simulated, routine line segment from IAD to JFK followed by a segment from JFK to LHR in which a mechanical problem was introduced that necessitated an engine shutdown and diversion from the original flight plan. The simulation included all normal communications, ATC services, weather, closed runways at the only favorable diversion airport, and later, an inoperative autopilot which further increased pilot workload. The scenarios were constructed in such a way that good crew coordination, cockpit communications, decision making, and planning skills were required, but they were not complex enough to preclude an entirely safe operation given proper performance and coordination.

This study allowed the examination of flight crew performance in a very controlled setting. Errors in performance were carefully monitored and recorded. Eighteen volunteer line crews flew the scenario and marked variations in the behavior of the crews were observed. Frequent problems were noted in areas related to communication, decision making, crew interaction, and integration. The presence or absence of strong leadership seemed to mediate the frequency and severity of the errors committed by the flight crews.

We took the cockpit voice recordings of these simulated flights and subjected them to a content-coding technique in which an attempt was made to classify each statement or phrase into categories of communication. Although the data I am about to present to you are based on a relatively small number of cases, some interesting relationships did emerge. Overall, there was a tendency for crews who did not perform as well to communicate less, but the type or quality of communication played a more important role. We found a negative correlation (r=.-51) between crew member observations about flight status and errors related to the operation of aircraft systems. In short, when more information was transferred about aspects of flight status, fewer errors appeared which were related to such problems as mishandling of engines, hydraulic, and fuel systems, the misreading and missetting of instruments, the failure to use ice protection, and so forth.

In a similar fashion, we found a negative relationship (r=-.61) between aircraft systems errors and acknowledgements to information which had been provided. In crews in which commands, inquiries, and observations were frequently acknowledged, these types of errors were less apparent. Acknowlegements were also related to fewer errors overall (r=-.68). It would appear that acknowledgements serve a very important function of validating that a certain piece of information has, in fact, been transferred. These types of communications also serve as important reinforcements to the input of other crew members. If you think about this relationship it appears very logical. When you make an attempt to communicate with someone and that person does not say anything, are you as likely to initiate further communication with that person? Aren't you more likely to communicate further if they respond even in the simplest fashion (e.g. yeah, uh-huh, etc.)?

Commands were associated with a lower incidence of flying errors (r=-.64): errors related to power settings, neglect of speed limits, altitude errors, and the lack of formal transfer of control between captain and first officer. Often communications of this type seem to assure the proper delegation of cockpit duties, but I would also like to suggest that too many communications of this type may have negative consequences. The use of commands provides a very good illustration of the interpersonal styles that I was referring to earlier. An identical piece of information can be related to other crew members in one of several different ways. For instance, a communication such as, "Check the plates for that profile descent procedure," which would constitute a command, could also be relayed, "I think we should check the plates for that profile descent procedure," an observation; or "Why don't we check the plates for that profile descent procedure?"-- an inquiry. Among other things, the overuse of commands is a good way to become one of those overbearing captains that I referred to earlier.

We also found some evidence for higher rates of response uncertainty, frustration or anger, embarrassment, and lower rates of agreement in crews who tended to make more errors. However, in these cases it is impossible to tell whether the communications gave rise to the errors or the errors gave rise to these types of communications.

In addition to the importance of communication style, the precision of communication plays a pivotal role. Unfortunately there is an abundance of stories where each pilot thought he or she knew what the other meant or intended to do when in reality they did not. The following report to ASRS conveys the potential severity of this state of affairs:

> I was monitoring an autopilot single autoland approach on a flight to Runway 22L at Newark in visual conditions when a GPWS "pullup" warning occurred...the entire crew's attention was directed toward confirming configuration, position, speed, and sink rate, all of which were normal. I commanded, and simultaneously with the copilot, selected 50 degree flaps, considering that the landing flaps not selected mode was possibly the reason for what I considered to be "probably" a false warning due to a possible failed flap position switch. The FE said something which I could not clearly grasp while

> the "pullup" was sounding. As the flaps extended to 50 degrees, the GPWS warning silenced. A normal auto landing occurred. Sounds simple and somewhat everyday--until during taxiing I was informed that the FE had inhibited the GPWS without being commanded to...he stated he asked me if I wanted it cancelled, but I did not reply so he assumed I did.

These observations about communications also point to a classic demonstration of a person-situation interaction. When looking at *individual* performance situations, we would expect that goal orientation is the important factor and that expressivity is generally irrelevant. Obviously, it is a safe bet that most operators prefer to have highly goal oriented people flying in their cockpits--so based on individual performance evaluations they select for this type. However, if the assumption is correct that a significant number of accidents are due to breakdowns in *crew coordination* and information transfer problems between crew members, we might wish to focus on a slightly different profile. The philosophy of redundancy built into the crew concept should be a testament to the need to look at the entire crew's performance. Thus, I would predict that the best crew performance would be associated with a situation where the captain, in particular, is high on both goal *and* group orientation. Theoretically, these individuals are more likely to be competent in both the problem solving, achievement aspect of performance and in the management of the human resources in the system. Traditionally, the group orientation dimension has been largely ignored.

It is an interesting fact that we occasionally hear complaints from industry sources about problems with "macho" pilots. It is probably not necessary for me to go into a detailed explanation of what I mean by that-- I think most everyone is familiar with the "scarf and goggles" type, and if not I will refer you to Tom Wolfe and *The Right Stuff* for a far more elegant explanation of the concept than I could ever hope to give. Research has unequivocably documented that the highly masculine individual, "macho" if you will, is very likely to be high in goal orientation and low in group orientation. Again, it would appear from a resource management, communications, and crew coordination standpoint, that this is not the best profile. Such a pattern may be very functional for astronauts and fighter pilots, but not so functional in today's air carrier operational environment. I have seen evidence of the "right stuff" mentality on several occasions while riding in cockpit jumpseats. On one such occasion, the first officer was flying a very complicated approach, and the situation was being compounded by other factors--controller-initiated rapid descent, moderate chop, a last minute change in the active runway; and during all this time he was trying to keep track of two other aircraft in front of him on short finals while the captain was laughing hysterically at his plight (and doing little else). When the first officer made some comment about not being very fond of his present circumstance, the captain simply informed him that the reason he did not like it was because he had been flying with too many "damned conservatives." I will submit to you that you might be just a little hesitant to point out to an indivdual such as that captain that he's "low and fast" at the outer marker.

Another area of concern these days is the socio-pscyhological impact of increasing automation and the more wide-spread use of direct air traffic control in increasingly dense traffic situations. There is a distinct danger that taking more and more functions out of the pilot's control, while decreasing workload, may create a psychological sense of loss, or diffusion, of responsibility (e.g. Darley & Latane, 1968). A number of investigations have documented that in such situations, humans are not as quick to take action in emergencies or they simply redefine potentially dangerous situations as non- threatening. The most widely cited example is the murder of Kitty Genovese, who was stabbed repeatedly in front of 38 witnesses, none of whom called the police. Now these witnesses were not terrible people. Their statements revealed that most thought that someone else would call or already had called for help. As automation takes over more of the actual flying responsibility, pilots may find themselves in similar states of mind. They may become hesitant to question an automated system--"this can't be wrong, the computer is flying this thing,"--and thus, not as likely to take action or label situations as potentially threatening. We may also see reduced vigilance on the part of crews even though they are consciously aware of their responsibility for the aircraft. As the role of the pilot evolves more into one of "systems monitor," aviation psychologists and training specialists must seek ways of dealing with this potential problem.

A related problem may be developing with respect to the air traffic control system. A NASA study using the data base from ASRS seemed to indicate that there was a greater risk of collision, as measured by incident reports, when aircraft were under direct control. There are several possible interpretations of these data, among them that aircraft are more likely to be under direct control in dense traffic situations and thus more susceptible to collision; but one possibility is suggested by the same line of socio-psychological research that I alluded to earlier--diffusion of responsibility. Again, we have a situation where everyone assumes that someone else is minding the store, and as a result, people tend to be less vigilant. One of my favorite quotes comes from a very senior captain who apparently feels that automation and direct radar control are already causing problems. He said, "We are the best trained instrument pilots in the world, but we are not trained to look out the window anymore. It's easy to go cross-country on radar, and have somebody else do everything for you. Sometimes we say to each other up there that the janitor could fly the plane as well as we do."

Another interesting issue is the socio-psychological dynamics of flight crew interaction in emergency situations. In the Ruffell Smith study, which I talked about earlier, it was found that the overload of a particular crew member was often associated with the commission of serious errors. It was also noted on several occasions that the captain tended to take over flying responsibility in the emergency situation. In one respect, this seems the proper approach and completely normal given the role structure in the cockpit, what captains are often told in training, and the fact that the captain is the one ultimately responsible for the aircraft and its passengers. However, from a psychological standpoint, this may not be the best strategy. It is probably true that the decision maker is the least appropriate person to be overburdened in such situations. In stressful circumstances, humans tend to exhibit a narrowing of perceptual attention such that it is perhaps not optimal to have the captain involved in flying the aircraft, coordinating the activities of the flight and cabin crew, and making the ultimate decisions regarding actions to be taken. One of most important jobs of the captain or any other manager is the effective delegation of responsibility. In the Ruffell Smith study, immediate benefits were derived by captains who elected to hand over flying of the aircraft to the first officer while making decisions about how to handle problems. In addition to allowing the captain more time to serve as an effective decision- maker, such a strategy allows other crew members a greater sense of responsibility.

This greater sense of responsibility among subordinate crew members is obviously important because of the strong probability of reduced vigilance on the part of other crew members if the captain apparently has "everything under control." This effect could easily be magnified if, for instance, the captain has relieved the first officer of flying responsibility after the onset of an abnormal situation. On the other hand, the captain must be careful to delegate responsibility effectively. We also saw situations in the Ruffell Smith study where captains over-delegated responsibilities to the point of causing overloads on the other crew members. In one instance, the flight engineer was interrupted so many times while trying to calculate how much fuel to dump that 77,000 lbs. too little was dumped, and the aircraft landed at a gross weight in excess of that required for the available stopping distance.

Coupled with all the other factors that can contribute to the lack of effective crew coordination in emergency situations is the fact that group members generally do tend to be more dependent on leaders under stressful circumstances. As a result, the captain is likely to bear even more responsibility for monitoring and orchestrating the actions of the crew. The type, and ramifications, of the leadership style exercised by the captain represents an excellent example of the person-situation interaction that I have been talking about today. For example, the authoritarian type of individual may be generally disliked as a captain in normal line operations. However, in emergency situations, such an individual would probably have no difficulty taking charge very effectively and may be well-suited to dealing with the added burden of leadership. On the negative side, this person may contribute to the sense of loss of responsibility among subordinate crew members. The democratic, socially concerned captain would most likely be extremely effective, or at least very well liked, in routine operations, but find it difficult to assume strong leadership when the circumstances call for it. Several airlines are developing resource management training programs recognizing that such factors are important to the safe and efficient operation of aircraft. Among other things, these programs are to provide feedback to pilots on certain personality dimensions and their role in crew coordination. These programs obviously do not seek to change personalities, but assume that an increased awareness of group dynamics will cause many to think before reacting in ways that are detrimental to group cohesion and performance.

There is yet another issue which we will be hearing more and more about and one which is bound to affect crew coordination in some way. That concerns the integration of female crew members into the previously all-male domain. Let me first say that all the evidence indicates that women pilots, in terms of individual performance, can be equal in every way to male pilots, despite the various remarks you hear out on the line about, "women falling apart under pressure or not being big and strong enough to kick the rudder when you really need to." These excuses are no more true than they are for some men. The real problem, however, is the fact that these attitudes persist. At present, we really do not see any serious difficulties in routine operations, but in work overload or emergency situations, especially where male crew members have reservations about the competence of female crew members, crew coordination could disentegrate. The recommendations and actions of female crew members may be questioned more often or not accepted. Male crew members may be more likely to take over some of the female's responsibilities causing further work overloads. I would like to stress again that the biggest problem will probably be in overcoming some male pilots' attitudes about the competency of women on the flight deck. I had an interview with a female first officer who felt that her sex had already influenced the captain's reactions in a dangerous situation. She was flying a leg where they received an engine fire alarm shortly after rotation. She reported that she scanned the gauges and found the engine still developing full power--making a conscious decision to let it go for a brief period of time until she could make sure the aircraft was stabilized in a positive rate of climb. However, the captain, without any warning, reached over and pulled the engine fire extinguisher bottle, killing the engine. As she said, "I almost lost the airplane." He then took over.

Now we really do not know enough about the situation to debate whether this captain made a right or wrong decision in that split second, and of course, he was the captain and it was his decision to make. Nevertheless, at that critical point, transferring control could have been disastrous--she was making the takeoff, she already had the feel of the aircraft. In her consultations with the chief pilot of her company, he tended to confirm her belief that the captain should have let her stay with it at least a while longer, and that there may not have been any real reason to kill the engine at that critical stage. Of course, hindsight is wonderful. It is also impossible to say that the same captain would not have done the same thing with a male first officer, but she very definitely felt because of the way he acted and talked later, that the major problem was her sex.

As women develop more seniority, we will obviously start seeing female captains. If these attitudes persist, the junior crew may be more likely to usurp or question her responsibilities. The previously alluded to "macho" pilot is likely to play a pivotal role in such scenarios.

In closing, I would like to ask the many aviation psychologists here today for assistance in researching the issues that I have tried to outline. We are faced with a big challenge in light of the rapid evolution of the aircraft and its systems and most importantly, the resulting change in the pilot's job. We have a safe system, but with more knowledge about the coordination of the human element, we can do a better job.

Invited address delivered to The Symposium on Aviation Psychology held at The Ohio State University, April 21-23, 1981.

REFERENCES

Darley, J., & Latane, B. Bystander Intervention in Emergencies: Diffusion of Responsibility. *Journal of Personality and Social Psychology*, *8*, 1968, 377-383.

Foushee, H. C., & Manos, K. L. Within Cockpit Communication Patterns and Flight Crew Performance: in Billings, C. E., & Cheaney, E. *Information Transfer Problems in the Aviation System*. NASA TP 1875, 1981.

Ruffell Smith, H. P. *A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance, and Decisions*. NASA TM 78482, 1979.

Spence, J. T., & Helmreich, R. L. *Masculinity and Femininity: Their Psychological Dimensions, Correlates and Antecedents*. University of Texas Press, 1978.

Wolfe, T. *The Right Stuff*. New York: Farrar, Straus, & Giroux, 1979.

MONITORS OF HUMAN PERFORMANCE

John W. Senders
University of Toronto

I tried to visit the Aero Med laboratory in order to find out what was going on. During that time, in order to get in, one had to apply for a position. The inevitable happened, they hired me. I had a degree in Psychology from Harvard, they employed me as an engineer to run the Apparatus Section of the Psychology Branch. I started as the resident engineer of the Branch and in about three weeks, it became clear to me that the people doing the research really didn't know what they were doing and I had some better ideas. I decided to start my own research program, and very rapidly discovered that psychology was easier than working and have continued in that field ever since.

Aviation Psychology is rather like the Agatha Christie play which has been running in London, England since 1955 and is the longest running stage play in history. The lines are always the same. Only the actors change. I look at the topics which are being talked about today and tomorrow, I see little difference from what was talked about in 1950, 52, 54: choose any year, you find the same play, the same topics, the same problems, regrettably, often the same approaches. These are modified by the advent of new technology. As we moved from ball-disc integrators that Con Kraft mentioned to analog computers to primitive stand alone digital systems involving vacuum tubes to the early solid state devices and now onto the micro processors, we see an application of the new technology to the old problem. We don't solve the problem. What we do is merely to restructure it. What was once a problem in the display and presentation of information has now become a problem in the "integration of digital control systems into the complex multi-dimensional cockpit management problem". It's still somebody flying an airplane from A to B on the basis of the information which is presented to him. We modify the information with what we have at hand. I feel that, for the most part, studies directed towards a solution of a particular problem have never actually solved them. Usually the problems are solved by the intelligence and knowledge of the research people who happen to be there at the time. In my view this characterizes what has been going on in aviation psychology as well. I noted with interest and surprise that Paul Fitts is quoted as having suggested in 1951 that the computers or machines should monitor men rather than the other way around. I know that I spontaneously arrived at that conclusion in 1952, as well.

I first visited The Ohio State University because Walt Grether didn't want to go one time when Paul Fitts had invited him, and he suggested I go instead. I asked where the lab was and was told to go to the Natatorium. I commented that I had had Latin but wasn't sure whether that was where you got born or went swimming, or put your behind. Walt thought for awhile and said, "Well I think it's probably all three as a matter of fact."

In any event, I went to the Natatorium and met Fitts and many of his co-workers and students. I never was a student of Paul's nor for that matter a colleague, except that we were both active in the same general research area. So both of us, and Paul before me, had suggested that mahcines might better watch men.

If one is to appreciate the whole notion of man-machine system analysis (MMSA), it is necessary that one get into the proper frame of mind. Figure 1 shows a system of a human operator (HO) and a working place (the plant). There must be assumed a task to be done. If there is none, then there is no system. Since the system will not be doing anything to any criterion, any performance of the system will not lead to changes or corrections in the activity of the HO nor will it make any difference whether HO does anything at all. (The laboratory exercises which engineering psychologists and man-machine systems engage in are more or less synthetic and the outputs of such experiments are suspect for that reason.)

A system to be analyzed has then HO, a plant, and some criterion of performance. Usually the plant is definable and quantifiable. Thus a power generating station can be characterized as a set of equations which state what each variable will do as time unfolds given a specified set of starting conditions and disturbances. Such calculations must assume, however, a law of control. A law might be that no control at all will be exercised: all the controllable variables will be left as they are at time 0. Or it might be that any or all of them will be allowed to change in accord with the equations (some may change and some may not). Or it might be that some specified kind of goal directed controller is assumed. The latter is more likely to be the case and it is to determine the form of this assumed controller when HO's are involved that MMSA is used. Of course, there are automatic controllers which depend on more or less sophisticated computers (from a thermostat's on/off 'computer', to a large digital computer which controls many variables). Even in these cases there is at some point in the system HO, doing something to some criterion (it is to be hoped).

It is assumed in Figure 1 that any automatic controller is subsumed into the plant so that the role of the HO is always that of receiving signals from the plant and delivering signals to the plant. These two interactive functions define the fields of control and display. Much of the early work in MMSA was on these two areas. The shapes of controls, their orientation, texture, force requirements and so on for an almost unlimited array of variables have been investigated in an almost unlimited number of experiments. Since many of the variables are not (apparently) capable of analytical formulation - no mathematical equations can be written to describe the shape of a knob for example -, the relationships between shape and the like and HO performance must be found by experiment. The same is true of displays. It is difficult to imagine a theory of displays which would allow prediction of performance as a function of the change in shape of the bezel surrounding a dial from circular to oval to

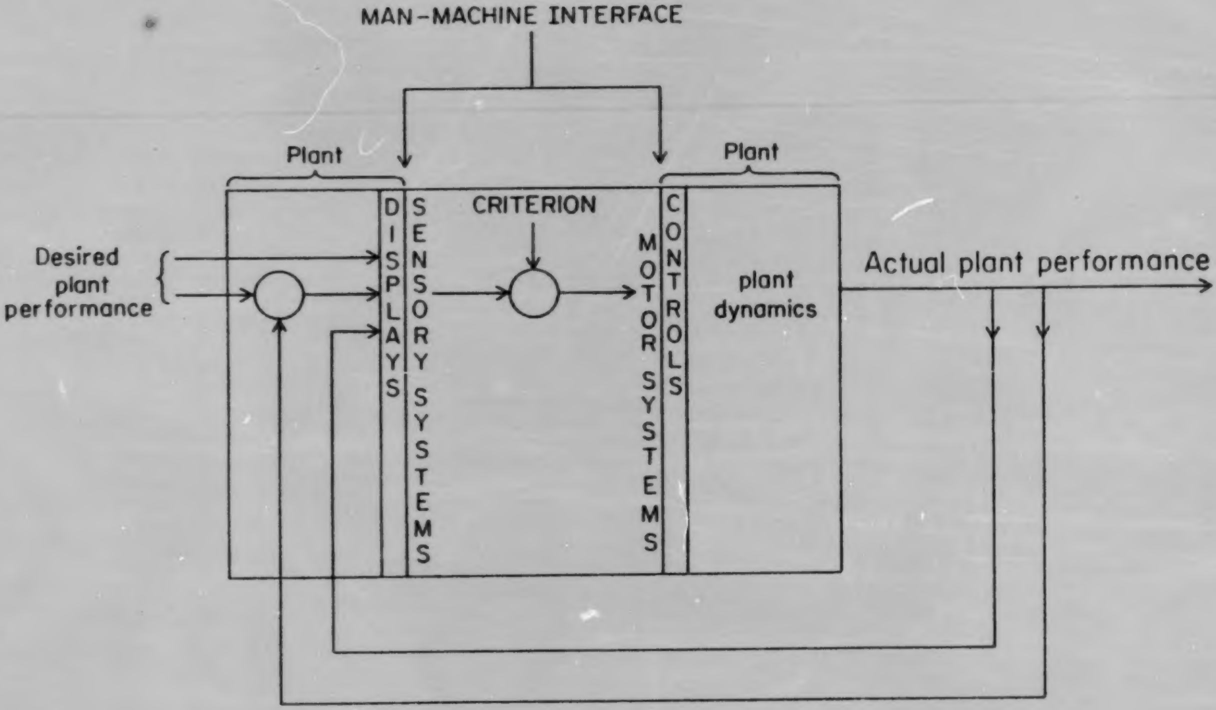

MAN-MACHINE INTERFACE
Plant
Plant
CRITERION
Desired plant performance
DISPLAYS
SENSORY SYSTEMS
MOTOR SYSTEMS
CONTROLS
plant dynamics
Actual plant performance

Figure 1. Conceptual model of man-machine system.

rectangular to square. Such things must be investigated. And of course, like controls, displays in all their aspects have been investigated at great length.

These aspects of MMSA are singularly a function of the psychophysics of perceptual processes and the physical anthropology of the hand, arm, foot and leg (and occasionally, these days, of eyeball and mouth). The laboratory work has been aimed at finding out as much as possible which is relevant to design about how the eye, the ear and just about everything else works so that there can be a match between the controls and the displays and the eyes and the hands (for example). Typically the visual acuity of the eye determines the minimum size of the indicator marks. The lower the light level, the larger must the mark be since acuity diminishes as light level diminishes; and so on throughout the whole range of sensory systems and variables. The results of such studies have been sorted, interpreted for engineers, and compiled into handbooks of 'human engineering design' (Van Cott and Kinkade, 1972).

The step from empirically determined functional relationship between design variables and human performance to quantitative analysis of the relationship between the dynamics and statistics of plants and the performance of HO's was a large and significant one. There is still much controversy about its validity. In order to make the difference clear and to explain why there is controversy, it is necessary to expand the simple diagram of Figure 1 into more detailed forms of Figure 2.

HO must be considered as a part of a closed loop system. This has not been generally true for most of the experimental work of engineering psychology and human factors. The behavior of HO had no effect on the next stimulus presented to HO. Whether an act was good or bad, right or wrong, the sequence of signals presented was predetermined (usually randomized). Thus the behavior was open loop as in Figure 2a. Engineering analysis, on the other hand, particularly that of control theory, was concerned with closed loop systems as shown in Figure 2b. Here the consequences of action are manifest in changes in signals presented to HO. This is, of course, much more like the normal kind of existence that human beings are engaged in all the time. What one does has an influence on what one does next. Although some engineering theories are not intrinsically closed loop in their formulations, information theory for example, the human being acting as an information transmitter usually alters his behavior as a function of the information he has transmitted. There are few memory less channels (other than the telegrapher of anecdote who received but was unaware of the news of Lincoln's assination, and the typist who is unaware of what he is typing). The fact of memory makes open loop infomation transmission relatively rare and for the most part unimportant in the world of MMSA.

Now we can look at Figure 2c. Here is HO coupled with a plant by means of controls and displays. The HO has internal feedback from sensors in his effectors and external feedback through vision of the positions of his arms and hands and so on. These feedbacks, which are not usually included in the description of the relationship between HO and the plant are nonetheless of great importance in HO's control of his own actions. Similarly, there are internal and external feedbacks which tell HO where

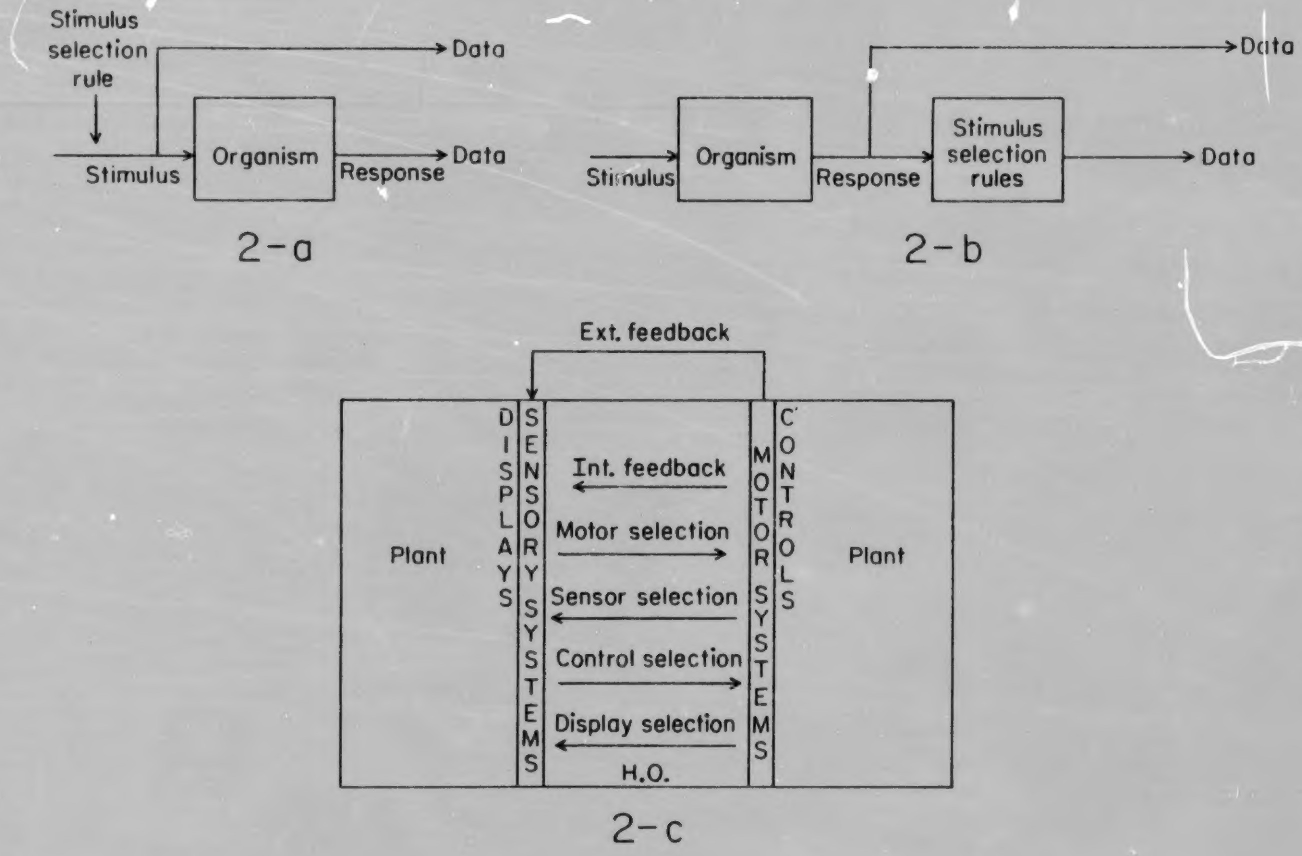

Stimulus selection rule
Stimulus
Organism
Response
Data
Data
2-a
Stimulus
Organism
Response
Stimulus selection rules
Data
Data
2-b
Ext. feedback
Plant
DISPLAYS
SENSORY SYSTEMS
Int. feedback
Motor selection
Sensor selection
Control selection
Display selection
H.O.
MOTOR SYSTEMS
CONTROLS
Plant
2-c

Figure 2. Detailed view of man-machine system.

his eyes are pointing and where his ears are listening. The major feedback, of course, is through the plant. The behavior of the plant connects the control to the display. If HO is to control the plant against instability or purturbation, he must choose a form of behavior which is appropriate to the dynamics of the plant. If the plant requires a left turn of the control to bring it from a low state (for some variable which needs to be controlled) to a higher state, the HO _must_ move the control to the _left._ Thus, much of HO's behavior is plant-controlled. In fact, it must be assumed that for a highly skilled HO, operating close to the limit of margin of his capacity, all his behavior is dictated by the nature of the plant. Free will does not exist under these conditions. If there were any freedom of choice by means of which the HO could improve his performance, he would exercise it and become controlled by the needs of the system. Under these conditions, when HO is controlled by the nature of the task and of the plant, we can apply the techniques of MMSA. This is so because of the availability of methods which allow us to compute the nature of a 'black box' given that we know its input and its output (and given that we can accept some limitations on the form of the expression that we obtain for the black box). This is the typical engineer's approach to engineering psychology. You have a human operator controlling the plant. The actual state of the plant is fed back and compared with the desired state. The difference between the two is the input of the operator. Of course, this must be conceived as a multi-dimensional task of many channels, each containing large quantities of information of different sorts.

The alternative (particularly with respect to aviation but also to a degree with respect to complex process control and the difference between the same nuclear powerplant operation and the manipulation of large aircraft is getting smaller and smaller all the time) is that you have a human operator and an auto pilot. The role of the human operator is to monitor the signal coming to him from the desired state and at will, switch the auto pilot in or out. The auto pilot can exercise the control parallel with the human operator; many of the early process controls were of this sort. The configuration is not unique to aircraft. They existed in power generations in the 40's and 50's and to some extent exist today in many chemical process control systems.

An alternative had the human operator engaged in continuous control activity process with the computer observing the behavior of the human operator by looking at the error signal and making on some criterion a decision to transfer control to itself. It is of great interest to contrast the two situations. In the one, the human operator monitors the behavior of the computer. The computer exercises immediate on-line control. It is a well known fact, well known to all of us in psychology and aviation psychology, that even if practice does not make perfect it at least tends in that direction. The old data from the cigar rolling industry showed that even after 25 to 30 years of rolling cigars, there was still an improvement. Paul Fitts' data taken on the simple task of sine wave tracking showed that if properly looked at, you get a constant increase in performance with no sign of leveling off. In other words constant practice leads to constant improvement. The notion that one can put the human being in the monitor task and expect that he can take over control in an emergency is a delusion. We have seen massive evidence of that at Three Mile Island and we might expect similar evidence if we were so unlucky as to have failures of the automatic systems that control modern aircraft.

The alternative is to put the human operator in the control loop and to keep him there so that in a sense, he is never out of practice. In its simplest form, the automated check pilot merely looks at the error signal and if the error exceeds some predetermined boundary conditions, it disconnects the human pilot and takes over control. I first suggested this to a technical group in 1952 at the Naval Training Devices Center at Sands Point, Long Island. At that time we were in the world of analog computers. It wasn't clear to every one that computers were that much more reliable than human beings, and it was generally felt then, and is still felt now, that the proper role of robots should be to relieve man of tasks that we would otherwise assign to him.

There is some justification for this. For example, the original purpose for having human being control machines and to watch them, was that the machines were relatively unreliable. I came in last night in a rather new Navaho, I was the only passenger; there were two pilots (this calm air) and I chatted with the copilot as we flew in. He was reassuring me about his skill. He said, "I used to be a University student, but I dropped out to be a pilot just two years ago, and here I am flying this airplane, and it is really great". That cheered me up a great deal and I asked him what his subject was. He said, "I was doing psychology". I felt very much at home. He was flying the airplane and to me it seemed that we were remarkably stable. "Are you on autopilot?" He said, "Oh no. Many of these planes don't have it, I don't know whether this one does or not. (to the pilot) "do we have an autopilot?" The pilot reached over, flipped a switch, and looked at some gauges and said, "Yes". We immediately went into a small amplititude continuous oscillation, which is characteristic of the behavior of the autopilot but not the behavior of the pilot himself.

In the earlier days, autopilots were even less reliable. Many a man has been killed by a failure of the automatic system, flying at low altitude on autopilot control. Now, however, the roles are fairly well reversed. We have at hand an enormous number and variety of cheap, highly reliable digital systems which are rapidly being incorporated into aircraft and into all other kinds of things that human beings deal with. The question that confronts the aviation psychologists is how best to act as both an optimizer of the man/machine system and as a preserver of the psychological functions of man. If man is relegated to supervisory and monitoring situations while the ever more competent robots of automatic control take over his functions, there is literally no end to that. You will, in due course, see human beings designed out of these systems, except possibly for window dressing; to satisfy regulations or to relieve public apprehension. And even that public apprehension will vanish as the new generation of people adapted from infancy to digital devices grow up and become dominant in this population.

It has been said that theories don't die, only theorists. It is also true that user populations are never converted, they merely fade away. What you get is a new user population which has different expectations. I foresee that if we don't make an effort to preserve the human function, it will, in fact, be eliminated because there will be general public acceptance of the idea that it should be. To all of us over the age of fifty this seems bizarre but to those under the age of twenty, it's not bizarre at all. It is, in fact, the inevitable conclusion of the large scale digital revolution that is going on now. What is the alternative? It might perhaps be to have a multi-function computer in a backup mode. Man should be in control at all times until it is demonstrated that he is unable to maintain control. This will preserve his capability of exercising control in the event of computer failure, and no matter how reliable our systems are, as they become more complex and more numerous the probability that at least one of them will fail becomes relatively large. It will also preserve his sense of personal value in the system.

In order that the man be kept in the system, he has to be maintained. This means that his behavior has to be compared with some standard. What I have in mind here, is that we have a model, not of an ideal man, but of an ideal version of the particular man. There is no necessity for absolutely perfect control of aircraft. There is always some permissible variance. The model should represent what we should want the human being to accomplish. The system would compare the behavior of the multi-dimensional model of the human operator with the actual output of the human operator and feed the difference between these two signals into a monitor. The function of the monitor is to determine the state of the human operator and to determine whether or not it is expedient to exercise control and to switch the human operator out of the system. The function of the machine is to monitor the reliability and performance capability of the human operator.

Any good check pilot should also be a tutor. The function of the tutor is to feed information based on a comparison of the model and the human operator back to the human opeator. Then the human operator will know how well he's doing and what he should do to improve. In order to do this, we don't have to go to the extent of solving the problem of artificial intelligence. If we did, of course, we would have another black box called AI, the ideal senior pilot which at every stage in the development of human pilots is doing its best to make sure that the human pilot is always at the peak of capacity. The proper function of the digital part of the system is to know what the human being does do, to know what the human being should do, to make an executive decision that the human being is within or without acceptable limits and to use the differential information between the do and the should do to modify behavior. In other words, to act as a teacher or tutor. Naturally, the system can introduce, if it wishes, tests into the operation of the machine. These can either be real or simulated. Real in the sense that the tutor part of the system could forceably cause the system to deviate in order to elicit particular kinds of behavior from the human pilot. It could drive the system to states within acceptable limits, which would ordinarily demand corrected behavior and examine that behavior. Organized programs of instruction could be incorporated into the function for which the person is employed. That is, of course, proposed not only as something to be done with an aircraft, but with any plant, in particular, any plant which involves large sums of money, great possibilities for mischief, and opportunities for destruction; nuclear power plants and super tankers are examples.

Many years ago at a conference on reliability, I had the temerity to suggest to a number of engineers that there might be some optimum, less than maximum, level of reliability. The reason for that would be that as reliability of systems increase, the opportunity for exercise of diagnostics and corrective action by the human being decreases and therefore the expected duration of down time of the system has actually increased. I do know that I forget how to repair, with modern reliable devices. With the older machines, which breakdown quite frequently, I am usually on top of the art. Although the total time down may be the same, the duration of down time on any one incident is less. It might be the case with complex systems like aircraft and nuclear power plants, that it would be better to have relative unreliability. A world famous pianist is said to have replied, when asked "why do you practice every day, after 50 years of performing?"; "if I fail to practice one day, I know it; if I fail to practice three days, the public knows it." Supreme skill is maintained only be continuous exercise of that skill.

Those of you who are pilots know that if you haven't flown for a bit of time, you are not really quite as sharp and don't quite have the style that you'd have if you had been flying every day. When we substitute mechanical devices for human beings, we are not in reality changing the organization of the cockpit. We are substituting limited artificial intelligences for the human beings who otherwise would occupy these posts. This is essentially technological unemployment. The uncontrolled rush to computerize everything and to have man "sit with folded hands" may turn out to be the biggest system design error of all.

REFERENCES

VanCott, H.P. and Kinkade, R.G. Human engineering guide to equipment design. Washington, D.C. Government Printing Office, 1972.

HUMAN FACTORS AND AVIATION SAFETY:
A Program of Research on Human Factors in Aviation

Stanley N. Roscoe
Behavioral Engineering Laboratory
New Mexico State University

ABSTRACT

Aviation safety is affected by human factors in system design, training, and operation, as it is by equipment factors, but in ways that are more complex and subtle and less well understood. Because of the complexities and subtleties, traditional approaches to the study of human contributions to flight safety and risk have tended to be subjective and piecemeal rather than objective and systematic. Research methods appropriate to the simultaneous study of the many interacting factors in the real world of pilots, controllers, and computers have been slow in coming. A new program of aviation research at New Mexico State University is being directed to these issues.

PROBLEM

Modern aircraft feature computer-aided communication, navigation, guidance, control, and display systems. Area navigation systems and control procedures have been implemented in a preliminary way and are being extended to provide vertical guidance, speed control, and energy management. Improved traffic control computers are being developed, a new microwave landing system (MLS) is being implemented, and a satellite global positioning system (GPS) and communication aids are waiting in orbit. These technological and operational advances will affect all types of flying; their benefits and demands will not be felt exclusively by the aeronautically sophisticated.

Predictably the situation just described involves complex changes in the roles of people and machines both on the ground and in the air. Understandably various elements of the aviation community are concerned about the long-standing human factors problems that are being elevated to critical levels and will surely get worse before they can be solved. However, before presenting a program of research on human factors in aviation, let us examine why such a program should be considered at all and what can reasonably be expected to result from its implementation.

HUMAN FACTORS AND AVIATION SAFETY
STANLEY N. ROSCOE

There appear to be two principal reasons for the growing feeling of urgency in developing a program of research on human factors in aviation. The first concerns the changing roles of flight crews and air traffic controllers with increasing computer-based automation and the impact of these changes on the people themselves--not only on pilots and controllers but also on passengers and support personnel. The second is the growing recognition among responsible people that both airborne and ground instrumentation, including simulators and other training devices as well as displays, controls, and communication equipment, frequently fail to provide necessary and sufficient information in a suitable form for current operations and that this need not and should not be the case for the future.

Function Allocation to People and Machines

Whether on the ground or in the air, some functions can be handled better by computers than by pilots and controllers, but the converse is also true. Nevertheless, the best ways to take advantage of the capabilities of each are not always evident and generally not clearly resolvable on the basis of current scientific knowledge. Furthermore, how well each can handle any given function depends greatly on how performance of the function is designed into the system--how computers are programed and how pilots and controllers are selected and trained and what types of displays and controls are provided to support their uniquely human abilities.

Other things being equal, as they seldom are, if people are to be most effective in complex system operations, they have to be kept busy. Humans are poor watch-keepers, or monitors, called on to perform only when something goes wrong or when the unexpected occurs. Computers, on the other hand, are excellent monitors and are capable of fast, accurate, and reliable responses in any situation that occurs predictably. The "Catch 22" is that the uniquely human capability to handle the unpredictable can be depended on only if the human is awake, alert, and ready to take effective action, and these conditions can be maintained only if the human is routinely involved and currently proficient.

Human Factors in Aviation System Design

The burden of human factors research in aviation is to provide a practical scientific basis for designing equipment and procedures and training and certification programs to optimize human performance of those functions assigned to pilots, controllers, and maintenance personnel. Whenever an airplane cockpit, a traffic control center, or an operational or maintenance procedure is designed, the designer has to make many decisions, whether consciously or otherwise, that will affect the performance of operational and/or maintenance personnel. Some relevant design principles have been established and embodied in minimum standards for certification, but these are generally not well stated, documented, or understood.

Human Factors in Aviation Training and Certification

Similar problems exist in the area of operator training, certification, and currency maintenance and assurance. Some principles of effective training and transfer of learning have been developed through research and operational experience. But once again these have not been well documented and are not well understood by many who are responsible for specifying the characteristics of training devices or for developing training programs. Clearly the major airlines have made the best use of advanced technology in training, but even here much improvement is possible and needed, and the benefits of their experience need to be passed along to the rest of the aviation community.

APPROACH

How should a program of research on human factors in aviation be organized and implemented to assure timely availability of workable solutions for the problems just described?

As a first suggestion, the problems can be approached in either a horizontal or a vertical fashion, and each has its place. By the horizontal approach we mean the development and validation of general principles of design for human effectiveness--principles that can be applied across the board whenever an operator is called on to perform a certain class of functions or tasks. By the vertical approach we mean the application, testing, and validation of horizontally derived principles during the advanced development of specific systems and prior to their operational certification. While this may cost time and money up front, it will surely pay off later.

As a general rule, horizontally oriented research tends to be done by universities and by a few small contract research groups. In contrast, the time and energies of research personnel in government laboratories and industry tend to be consumed by projects of a more typically vertical nature. The research programs at the University of Illinois on principles of display frequency separation, flight path prediction, and visual time compression are representative of the former type. The cooperative FAA/NASA programs on terminal configured vehicles (TCV) and the cockpit display of traffic information (CDTI) are recent examples of the latter type.

Display and Control Design Principles

Recurring problems in instrument design stem from the fact that whenever any particular function has to be implemented the designer has to make a number of decisions; he may or may not be aware that he is making decisions, and very frequently he fails to consider that the same alternatives have been dealt with many times before over many other drawing boards. Few laboratory directors, project engineers, training managers, or pilots realize just how many important design decisions are made in precisely this way. Nevertheless, this process has gone on and on throughout the history of aviation system development.

What are the sorts of decisions made over and over by different designers at their drawing boards? A few examples and some of the alternatives involved are listed below:

1. Size, scale factor, and sensitivity of a display
2. Direction of sensing: fly-to, fly-from, or frequency-separated
3. Visibility and reachability
4. Combinations of indications within a display
5. Display modes: alphanumeric, symbolic, pictorial
6. Arrangement of controls and displays within a panel or console
7. Feel of controls: damping, detents, feedback
8. Coding and function of switches, knobs, levers
9. Grouping of functionally related operations
10. Logic and coding of caution and warning indications

During the less than half-century since human factors engineering was recognized as at least a semiscientific discipline, countless horizontal and vertical research programs have dealt with such issues as those embodied in the list above. Nevertheless, different decisions have been made by different designers regarding similar applications of each of the items listed. Possibly because individual applications differ in subtle ways, the proper selection among design alternatives is not always evident even to the most experienced people in the field.

None of the required decisions would be particularly difficult to make if experts could agree on the correct choice among alternatives in each case or if there were available a sufficient body of objective data describing the consequences of any decision. It is a fact, however, that the experts do not agree. Some like the moving card, others like the moving pointer; some believe in "symbolic" others in "pictorial" displays, and so on. On the other hand, there is experimental evidence on many of these issues, but it is not complete and, in addition, lacks generality. When new problems arise that are somewhat different from the old ones that have been solved experimentally, it is not certain that the old solutions are applicable.

Solving each new problem or each new version of an old problem by experiment is simply not feasible. There is neither enough time, money, nor manpower to accomplish such a program. Nor is it satisfactory, in the absence of experimental evidence or unanimous opinion, to be confronted with the necessity for making what often appear to be arbitrary decisions. Often this necessity is avoided by authorizing development of several alternative versions of the same system in the hope that one will prove satisfactory. When this is done the designer knows in advance that a large proportion of his money is necessarily being wasted.

The hope that a largely horizontal program of research might ultimately reduce the designer's uncertainty appears to follow as a natural consequence of the present dilemma. A horizontal program, as a complement to existing vertical programs, carries with it the notion of generality of results, and this is what is needed. The horizontal approach implies in effect: let us not be totally diverted by the particular problems that arise from day-to-day, but let us consider the problem as a whole and attempt to arrive at general rules for displays and controls that can be applied successfully in any subsequent instance.

Training and Transfer Principles

Human factors problems associated with the training, certification, and refreshment of pilots, controllers, and support personnel have much in common with those encountered in equipment and procedures design, but there are also notable differences. In common is the situation that much of what is known and can be stated as principles is not necessarily known to the people responsible for operational applications. In contrast, however, this is not so much a problem for research as it is a challenge to spread the word to managers, administrators, and individual operators, including instructional system developers and professional instructors.

For example, the potential effectiveness of flight simulators in pilot training and certification is well documented, and in the case of airline operations, widely and legally accepted. However, simulators do not command similar respect and use in general aviation, air taxi, and commuter operations. Admittedly there is less economic pressure to replace flight training in less expensive airplanes, but the factors contributing to the relatively ineffective use of simulators in primary and intermediate training phases are complicated and subtle. To be cost effective, simulators must save their operators money by costing less to own and operate than the flight time they replace.

Possibly because of the outstanding success of airlines in using complex and costly flight simulators for training, the belief is widely held that simulators have to look, feel, move, and smell like airplanes to be effective. In a subtle way the airlines have been caught in their own trap. To persuade their professional pilots to accept the complete substitution of simulators for airplanes in the training and certification process, they have emphasized the total fidelity of simulators to their counterpart airplanes. The pilots, in turn, have so embraced the notion that a simulator has to be a tethered airplane that they are now insisting on simulators of higher and higher apparent fidelity.

This circular sequence of events and positions appears to offer mixed blessings. Clearly the importance of certain types of simulator fidelity has been well established both through research and operational experience, and this conclusion is gaining wide acceptance.

41

Unfortunately it is also evident that efforts to achieve ultimate apparent fidelity of simulators can be counterproductive. Not only is the cost far out of line with any possible benefits, but also the training effectiveness of such devices can suffer. Research has shown that certain intentional departures from literal duplication of aircraft characteristics can make possible training strategies far more effective than those currently employed.

Evidence for these strong and, to some people, heretical statements can be found in research on augmented feedback in training, on unrealistically exaggerated response lags and instabilities, on intentionally reduced visual cues in contact flight training and elevated workloads creating larger than lifelike stresses analagous to swinging a leaded bat before stepping to the plate. Similarly the unwarranted emphasis on ultimate apparent fidelity tends to discourage development and imaginative use of simpler and more flexible and reliable part-task devices and computer-based teaching scenarios that can yield even more effective training at a greatly reduced cost.

Reasons for the current state of our aviation training technology are not hard to discover. While the Department of Defense has invested vast sums in training-research simulators, virtually all of the research has been of a vertical rather than a horizontal nature. Because transfer of training experiments are difficult to conduct and also very expensive, such experiments typically involve comparison of two, three, or four training conditions treated as qualitative factors because they are actually composites of quantitative factors too numerous and confounded to unravel and manipulate individually. This approach is essentially vertical in that total simulator configurations are developed and then comparatively evaluated.

Results of such comparisons lack generality of application because they reflect only the combined effects of particular sets of values of the many component variables individually important in simulator design and use. To get at the main effects and interactions, statistically speaking, of the many independent design and use variables, a different research strategy is called for, one that is essentially horizontal rather than vertical. Fortunately a research paradigm new to the aviation community, but long used in the chemical industry, has been advanced by Dr. Charles Simon.

The practicality of applying this innovative research strategy to human operator performance and training is no longer a matter for speculation. More than half a dozen experiments conducted at the University of Illinois have involved experimental designs and multiple regression analyses of the type advanced by Simon. Also such a design was employed successfully at NASA-Ames Research Center in a study of pilot judgment of projected touchdown points on simulated landing approaches by reference to computer-generated visual displays.

Even more directly applicable, a transfer of training experiment recently completed at New Mexico State University included five simulator design variables, one training variable, and three transfer-vehicle configurations. The experiment, completed in less than a month, involved only 80 trainees, 48 of whom received training in individually unique simulator configurations. The experiment yielded reliable and unbiased regression equations for the main effects and first-order interactions of the six experimental variables for each of the three transfer-vehicle configurations. The specific findings of this experiment, dealing with a simple lateral-steering task, have little direct application to aviation but demonstrate that meaningful multifactor transfer experiments can be conducted effectively and economically.

FRINGE BENEFITS

Benefits of the NMSU aviation research program are not limited to the application of research findings and technological advances, although these can be expected to be substantial. The functions of a university are to educate as well as discover, and the production of scientists and engineers who specialize in solving human problems encountered in aviation system design, training, operation carries a high priority at this time; individuals formally trained and with research experience in these areas are in extremely short supply and are badly needed by the aviation community.

FITTS' PRINCIPLES STILL APPLICABLE: COMPUTER MONITORING OF FIGHTER AIRCRAFT EMERGENCIES

DR. JOHN REISING

Flight Dynamics Laboratory
Wright Patterson Air Force Base, Ohio

and

DR. LLOYD HITCHCOCK

FAA Technical Center
Atlantic City, New Jersey

ABSTRACT

When examining the impact of automation in the cockpit, a careful distinction must be drawn between the role of a multi-crew, civilian aircraft and a single-seat military aircraft. In the case of the former, there are some concerns that automation can have a negative impact on the crew by reducing their proficiency. However, in the case of a single-seat military aircraft on an attack mission, the role of automation is not only helpful to the pilot, but may be necessary to achieve mission success in the future. These distinctions must constantly be brought to mind in explaining the concepts of Fitts' principles. In the case of single seat military aircraft, it is believed that these principles still apply, and that the computer monitoring of system failures will further enhance their application. Three approaches to computer-augmented system monitoring are discussed.

INTRODUCTION

The increased capabilities of enemy air defense networks require that fighter/attack aircraft operate at low altitudes to take advantage of terrain masking. At the same time the requirement to perform the fighter/attack mission in adverse weather has increased and will be necessary in the event of all-out conflict. When these factors are combined, the mission of the single seat aircraft will become very demanding, and unlike the multi-crew, civilian aircraft, where concern has been expressed relative to loss of crew proficiency due to automation (Currey and Wiener, 1980), it will be essential to automate in order to reduce the pilot's workload to an acceptable level. In order to perform this mission successfully, the workload must be distributed appropriately between the pilot and the computer so that the pilot can do the things that he is most capable of, and the computer can perform the tasks it is best suited for. For example, the pilot is superior to the computer in the area of complex decision making and in his ability to interpret ill-defined patterns such as those which appear on ground mapping radar displays. The computer, on the other hand, has its strengths in its ability to store a great deal of information and to monitor systems very rapidly, e.g., 30 times a second. Proper allocation of these functions between human and machine is crucial to the successful completion of the fighter/attack mission of the future.

In this allocation process, consideration should be given to Fitts' principles: 1) that the tasks for the human should be intrinsically motivating 2) they should provide activity and 3) machines should monitor the humans, not the converse. There are two aspects to the third principle. The monitoring of humans by machines is showing great advances through the use of biocybernetics (Reising, 1979); however, this aspect will not be treated further in this paper. The second aspect, freeing humans from the monitoring of machines, is the main subject of the paper, and the key to solving this problem is through the use of the computer. Specifically, the computer should be given routine systems' monitoring and relatively simple decision functions. This allocation frees the pilot to concentrate on such mission related tasks as interpreting sensors and mentally previewing and rehearsing target area maneuvers -- tasks which provide both physical and mental activity, and are certainly motivating to the pilot. Fitts' first two principles, therefore, are basically guaranteed to be applicable based on mission-related tasks, and are not discussed in detail. Automation applied to system's monitoring will receive the main focus of attention. Computer monitoring of systems is employed in aircraft such as the F-18 and F-16. However, a more challenging aspect is diagnosing system failures, and structuring computer-aided emergency procedures.

The purpose of this paper is to describe three alerting systems with increasing levels of computer augmentation. The first approach uses a three tiered, computer-based warning/caution/advisory (WCA) system; the second approach expands the conventional three-tiered system to a five-tiered, tailored concept; and the third approach discusses artificial intelligence to illustrate future trends in this area.

APPROACH I: COMPUTER BASED THREE-TIERED ALERTING SYTEM

In a recently completed study at the Flight Dynamics Laboratory, the detection of emergencies relating to engine failure was examined. A computer-augmented presentation on a cathode ray tube (CRT) was used, as well as electro-mechanical instruments coupled with a WCA panel. The pilot flew a simulated mission composed of three segments: takeoff/climb, cruise, and weapon delivery. Randomly interspersed during the flight were a series of failures. These failures consisted of engine and hydraulic parameters going out of tolerance (either high, low, or fluctuation). The task of the pilot was to identify the failure and to report the parameter and its current value.

In the CRT version of the engine display, the indication of a parameter exceeding its normal bounds was shown by a bar going beyond either of two dashed lines which indicated the normal operating range of the

45

parameter (Figure 1). The indication of a fluctuation condition was shown by the rapid movement of the bar which had the word FLUX printed in the center of it. For the electro-mechanical instruments the indication of the parameter exceeding its normal limits was the same as that used in current cockpits and was achieved by the needle moving past certain areas of the gauge which were marked with green tape. In the fluctuation condition, the excessive state was shown by rapid needle movement. The results of this study showed that performance with the computer-monitored system/CRT presentation was superior ($p < 0.01$). The pilots could not only identify the engine failures in 40 percent less time but also could identify the failed parameter more accurately with the computer/CRT presentation than with the electro-mechanical instruments and WCA panel.

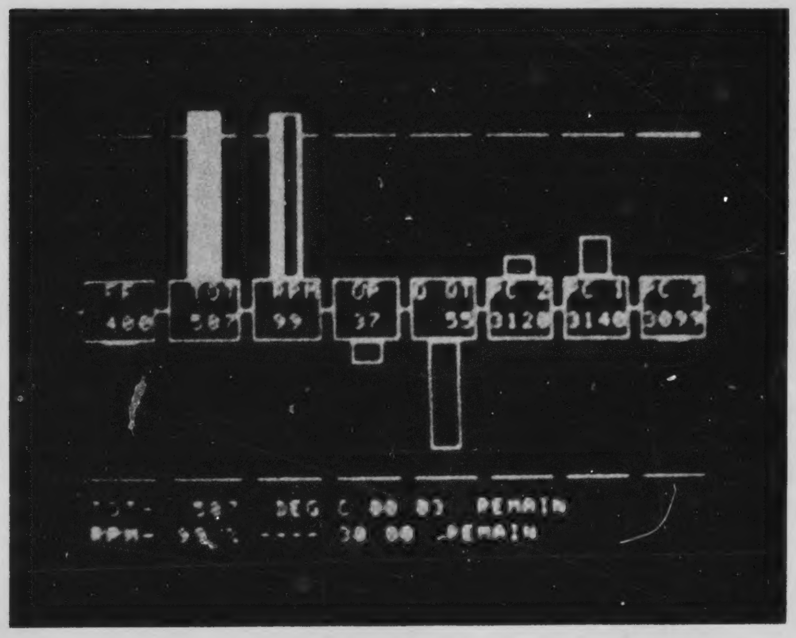


Figure 1. System status display on CRT.

A new capability also was exercised in this study. When a given parameter entered an advisory zone, i.e., a zone in which operation is permitted for a certain time period only, the bar continuously changed as a function of the time remaining in the advisory zone, i.e., the open bar started to fill in from left to right, rasterline by rasterline, in proportion to the time remaining. An illustration of this feature is shown for the RPM parameter in Figure 1 which has been in its advisory zone for approximately one half of the allotted time period. This indication is not possible with the electro-mechanical instruments. With computer augmentation, the probability of pilots having engine damage is greatly reduced because of the early and clear indication of the advisory status.

The overall conclusion derived from this study is that computer monitoring of engines, coupled with its natural display partner the CRT, not only provides a better means of indicating engine failures, but also provides a method of reducing potential engine problems through advisory zone indications. This study also expanded the definition of the advisory portion of the WCA System through the inclusion of the advisory zone concept. The next approach concentrates on the expansion of the caution and warning portions of the WCA system through the use of five-tiered, tailored concept.

APPROACH IIa: COMPUTER BASED FIVE-TIERED ALERTING SYSTEMS

In the course of reviewing the cockpit alerting system scheduled for incorporation into the Navy's F/A-18, it became clear that the long standing trilogy of Warning/Caution/Advisory was not adequate to meet the alerting requirements. This review had been initiated in response to the Navy's preliminary evaluation of pilots' concern over the frequency of occurrence of the caution tone in the F-18, often under circumstances where that level of alert urgency was questionable. This problem has been identified in commercial aircraft by Veitengruber, Boucek, and Smith (1977) and Cooper (1977). Out of the resultant discussions came the foundation of the alerting schema to be presented here. This approach was not incorporated into the F-18; a more immediately applicable fix was selected. However, it is presented here as a stimulus to more comprehensive considerations of the alerting systems for future aircraft.

The basic concept which underlies the recommendation to modify the currently accepted alerting system is the compelling recognition that the levels of urgency exceed the traditional three; in short, it became clear that there are Warnings, and there are WARNINGS. MIL-STD-411 defines a "warning" as a "hazardous condition requiring immediate corrective action". When the range of potential interpretations of the terms "hazardous" and "immediate" is considered, it is obvious that the scope of the basic definition is broad indeed. Similarly, a "caution" is defined by the same source as an "impending dangerous condition requiring attention but not necessarily immediate action", which proves to be an even broader categorization. In an attempt to increase the specificity, and hence the utility, of the alerting process, the basic three-category system was expanded to include five levels of criticality: two levels of Warnings, two levels of Cautions, and one level of Advisories. The definitions associated with the categories are:

W1 -- This is a true WARNING. It connotes a condition which threatens both aircraft and pilot with imminent catastrophic destruction. It must be handled "NOW" no matter what the circumstances. These conditions are so threatening that they should bring the pilot's attention inside the cockpit even during low-level, high-speed penetration and/or cause him to ignore a bad guy on his six.

W2 -- The conditions included in this category merit designation as "Warnings" because of the severity of their consequences if ignored. However, they fall short of the W1 category because the need to respond is urgent but not immediate; there is at least a short period (30 seconds or more) in which the pilot can attend to the pressures of his current flight environment before turning his attention to this problem.

C1 -- The consequences of the failures indicated by this alert are severe but not necessarily life threatening in and of themselves. If ignored, the subsequent failures might seriously degrade system capability at a critical phase of the mission and/or result in extremely costly damage to the aircraft. They could, under certain circumstances, develop into full blown threats to safety.

C2 -- Cautions in this category include abnormal system states or out-of-tolerance conditions which have implications for aircraft health and/or safety. They include conditions which, though not notably serious in themselves, would, if left unattended, lead to more significant/costly system failures. Also included would be failures which impact system capability under certain mission phases. Attempting to fly such phases with a failure degraded system would unnecessarily expose the aircraft and pilot to enemy counter action with no hope of mission completion. These are "need to know" items.

A -- The advisory category includes information regarding system state or mode/option selection which is of significant interest to the pilot but which does not, in itself, imply an unsafe or inappropriate condition.

These various categories could be communicated to the pilot as shown in Figure 2. The combination of visual signals, tones, and voice in a WCA scheme is also discussed by Seminara (1965). The use of a warning tone for W1 was derived from numerous pilot reports that the tone, at least the "tweedle-tweedle" used in the F/A-18 was far more compelling than the voice and would probably pull their heads inside the cockpit no matter how it was used. Since the "head in" response is the desired result of a W1 alert in order to look at the dedicated warning displays, it is recommended that it be coupled to the tone. The voice alerts were tied to the W2 category to inform the pilot of the problem facing him without pulling him back into the cockpit; allowing him to set his own pace in compensating for the failure based upon his assessment of the alert urgency and his present flight environment. The voice alerting was extended to the C1 level to emphasize this category more strongly

LEVEL	MASTER LIGHTS		TONE	VOICE	ANNUNCIATOR PANELS		
	WARNING [RED]	CAUTION [YELLOW]			CAUTION [YELLOW]	ADVISORY [GREEN]	MPD
W1	▨		▨				
W2	▨			▨			
C1		▨		▨	▨		▨
C2		▨			▨		▨
A						▨	▨

Figure 2. Possible five-tiered warning/caution/advisory system.

than the use of the caution lights alone. This permits the assured identification of the more critical cautions without tying a tone to all caution indications. W2 Warnings using voice would be phrased "Warning - Warning - APU Accumulator Failure". C1 Cautions would be similarly labeled as cautions to differentiate the two.

Though Figure 2 shows an annunciator panel to specify both cautions and advisories, the introduction of multipurpose programmable display (MPD) units into the modern cockpit provides an alternate display medium for these designations. However, it would be advisable to maintain the caution panel, even when using an MPD for primary presentation, to serve as a back-up designator in the event of major power failure. Similarly, the W1 dedicated displays should always be hard-wired through an emergency power supply no matter what the primary bus/display system used. For this reason the MPD is not shown as appropriate for either W1 or W2 conditions; however, the MPD could present additional data to be considered at a later time period.

APPROACH IIb: TAILORING EMERGENCY INFORMATION

In the recent past, once a condition had been labeled as advisory, caution, or warning it retained that designation under all circumstances throughout the life of the aircraft. The advent of on-board computers and their control of information through digital bus systems has eliminated that inflexibility forever. The designer of the alerting system can now tailor the presentation of information about a single fault condition in a way which reflects the importance of that condition in terms of the immediate operational environment. This approach was envisioned by Vanderschraaf (1976) in his Phase Adaptive Warning System (PAWS). For example, the fact that the canopy is unlocked could be displayed to the pilot in a number of ways depending upon where he is in his mission. The designated alert level could vary as follows:

A -- The pilot has energized master power, weight-on-wheels, engines at idle or below. No need to hit him with a caution light and/or tone while he is still at the line going through his start up sequence.

C2 -- Anytime canopy is closed, lock not engaged.

C1 -- Weight-on-wheels, engines above idle RPM to indicate taxi with unlocked canopy.

W2 -- Weight-on-wheels and engines above 90% RPM to indicate imminent take-off with canopy unlocked.

W1 -- Anytime canopy is unlocked and there is no weight-on-wheels indication.

Similar tailoring could reflect time available for correction and/or impact upon mission mode. For example, a failure of the radar cooling system which allows the pilot 90 seconds to secure the radar before it overheats could be given a C2 when it first occurs and transitioned to a C1 when it has not been corrected by securing the radar and only 30 seconds to secure remain, as long as the aircraft is in a benign operational environment such as the navigation mode. However, if the aircraft is in the air-to-air mode, the radar is locked on, the Sparrow missiles are armed, and the radar range is 10 miles or less (or some other appropriate combination of conditions), the failure of the radar cooling and its attendant implication for the imminent loss of that sensor system might merit a W2 as soon as detected based upon the advisability of continuing to prosecute the selected attack.

The purpose of this discussion of alerting taxonomy is not to present a five level system as being anymore sacrosanct than the preceding three tiered schema. Rather, it is intended to use the existence of newly developed systems technology to encourage a complete reanalysis of the approach to pilot alerting. Even with the augmented levels, it should be expected that there will be some animated discussions about whether a given condition should be labeled W2, C1, or C2; the extremes, W1 and A, are usually subject to greater inter-observer agreement.

APPROACH III: ARTIFICIAL INTELLIGENCE AND COCKPIT EMERGENCIES

The previous section expanded the concept of WCA into a five-tiered alerting system. The section also introduced the concept of tailoring emergency procedures to the particular state of the aircraft. This section will build upon the previous one and introduce the concept of using artificial intelligence to determine actions for the pilot and to reduce the information processing load placed upon him.

Artificial intelligence (AI) suffers currently from a definitional problem in that some view it as the construction of intelligent systems and others view it as a means of emulating the way human beings solve problems. The focus of the first definition is on the results or output of the system, while the second concentrates on the process by which the output is formulated. That issue is beyond the scope of this paper and the answer probably lies in a blend of the two. A quote from Hofstadter, (1980, p. 27) may help to clarify the role of AI and emergency procedures,

> The flexibility of intelligence comes from the enormous number of different rules, and levels of rules. The reason that so many rules on so many different levels must exist is that in life, a creature is faced with millions of situations of completely different types. In some situations, there are stereotyped responses which require 'just plain' rules. Some situations are mixtures of sterotyped situations -- thus they require rules for deciding which of the 'just plain' rules to apply. Some situations cannot be classified -- thus there must be existing rules for inventing new rules . . . and on and on.

The current types of computer-augmented cockpit alerting systems are operating at the stage of sterotyped responses with just plain rules. The next step is to go to Hofstadter's second level. That is, trying to determine some rules for deciding when to apply the just plain rules. In order to achieve this second level, a study is being conducted to examine existing AI algorithms and to see if they can be applied to aircraft emergencies. The third situation of inventing new rules is a step beyond the current capabilities.

There exists a series of AI programs which are used in medical diagnosis, e.g. MYCIN (Shortliffe, 1976). They evaluate a set of symptoms and laboratory reports and, through various crosschecks of just plain rules, determine on a probabilistic or heuristic basis which candidate disorder is most likely. It appears logical that one can look at an aircraft with emergencies as ill -- just as a human with symptoms is ill. We are examining the feasibility of applying the expert system concept to the diagnosis and resolution of aircraft emergencies. Conceptually there does not appear to be any problem; however, the practical aspects of interacting in real time with aircraft systems are currently being examined. If AI can be used to recommend and perform at least some of the pilot's emergency procedures, this will greatly reduce his monitoring problems. If some of the problems are of a lower priority, the software can recognize them and not reveal them to the pilot until he is in a less demanding mission segment. If some solutions can be implemented by the computer, the pilot need not even be told until the mission is completed. Finally, if a problem is very complex, AI can suggest solutions, indicate the probability of success for each, and can execute the selected candidate if the pilot gives a consent.

A key feature of these approaches is to apply Fitts' third principle by offloading systems' monitoring from the pilot to the computer and by giving him a much clearer presentation of information; the motivation and activity aspects of Fitts' principles would be guaranteed as a result of the handling of emergencies in the computer-based alerting system. This, in turn, will allow the pilot to concentrate on more demanding mission tasks. Moreover, through the application of AI, the second aspect Fitts' third principle will be eminently fulfilled. It will no longer be true that humans monitor machines, but rather that machines will monitor machines and make intelligent decisions. Through this helper, an R2D2 analogue, the dull, routine and boring monitoring of machines will be removed from the pilot, and a smart machine will free him for the much more challenging and complex tasks of which he is capable.

CONCLUSIONS

1. Computer augmentation of current warning/caution/advisory systems results in improved pilot performance in single-seat military aircraft applications.
2. Consideration should be given to expanding the three-tiered warning/caution/advisory system to a five-tiered, tailored concept.
3. Artificial intelligence may free the pilot from routine monitoring tasks, while allowing him to concentrate on complex and challenging tasks for which he is best suited.
4. Fitts' principles are still applicable in the single-seat, military aircraft. Computer monitoring of aircraft systems, coupled with intelligent machine decisions when emergencies occur, will insure that Fitts' third principle is still applicable and will allow the pilot to concentrate on tasks in which Fitts' first two principles (human motivation and activity) may be used.

REFERENCES

Cooper, G.E. A survey of the status of and philosophies relating to cockpit warning systems. Report No. NASA-CR-152071, NASA Ames Research Center, Moffett Field, California, 1977.

Curry, R.E. and Wiener, E.L. Some human factors aspects of cockpit automation. Proceedings from IEEE/AESS Symposium: Human Machine Interface in Airborne Systems, 3 December, 1980.

Hofstadter, D.R. Godel, Escher, Bach: an eternal golden braid. New York: Vintage Books, 1979.

Reising, J.M. The crew adaptive cockpit: Firefox, here we come. Proceeding of the Third Digital Avionics Conferences, 6-8, November, 1979.

Seminara, J.L. How to apply human factors principles in warning systems design. Machine Design, 1965, 37(23), 106-116.

Shortliffe, E.H. Computer-based medical consultations: MYCIN. New York: American Elsevier, 1976.

Vanderschraaf, A. Problem area: warning systems, Fokker, VFW aircraft. Proceedings from the 20th International Air Safety Seminar of the Flight Safety Foundation. Anaheim, California, October 25-29, 1976.

Veitengruber, J.E., Boucek, G.P., and Smith, W.D., Aircraft alerting systems criteria study, Volume I: Collation and analysis of aircraft alerting systems data. FAA Report, FAA-RD-76-222, May, 1977.

THE PERFORMANCE OF WARNING SYSTEMS IN AVOIDING CONTROLLED-FLIGHT-INTO-TERRAIN (CFIT) ACCIDENTS

JAMES P. LOOMIS* AND RICHARD F. PORTER**

Battelle
Columbus Laboratories
Columbus, Ohio

ABSTRACT

This paper examines the performance of two systems to prevent Controlled-Flight-Into-Terrain accidents, including their development and preimplementation issues and attitudes. The airborne version, the Ground Proximity Warning System, was required for certain large turbine-powered airplanes. The ground-based system, the Minimum Safe Altitude Warning, is a feature of the ARTS-3 system. Accident data from National Transportation Safety Board (NTSB) and reports from the Aviation Safety Reporting System (ASRS) were used in assessing performance. It is concluded that these systems have dramatically reduced accidents. Although false and nuisance alarms continue, no evidence suggests that they have caused any accident. The tenacity of the alarms--especially the GPWS--as well as appropriate triggering criteria seem to be basic to their success.

INTRODUCTION

Battelle's Columbus Laboratories operates the Aviation Safety Reporting System (ASRS) for the NASA Ames Research Center. The authors recently studied Controlled-Flight-Toward-Terrain (CFTT) occurrences recorded within the ASRS data base. This data base currently contains 24,000 reports of discrepancies and deficiencies in the aviation system, authored by pilots, air traffic controllers, and others.

The CFTT study disclosed interesting facts concerning the impact of two warning systems on the day-to-day operational aviation system. The systems implemented to preclude Controlled-Flight-Into-Terrain (CFIT) accidents are the airborne system known as the Ground Proximity Warning System (GPWS) and the ground-based system, known as Minimum Safe Altitude Warning (MSAW)--a functional element of the automated radar terminal system (ARTS-3) installed at some 63 major US airports.

The reports studied include valid alarms (some appear to have prevented a disastrous accident) and false or nuisance alarms (some of which created new hazards). It was recognized that if this information were supplemented with CFIT accident data for the periods before and after GPWS/MSAW implementation, plus information on the

---

*Manager, Transportation Systems Section
**Principal Research Engineer, Space Systems and Applications Section

attitudes and issues preceding implementation, an interesting and potentially useful paper might result.

WARNING SYSTEM IMPLEMENTATION

In December 1974, the Federal Aviation Administration (FAA) adopted a rule requiring a GPWS in certain large turbine-powered (turboprop and turbojet) aircraft by December 1, 1975 (FAA, 1974a). Specifically, the rule applied to turbine-powered airplanes with maximum gross weights over 12,500 pounds operated under Parts 121, 123, and 135 of the Federal Aviation Regulations. These parts covered all airlines certificated by the Civil Aeronautics Board, scheduled commuter carriers, and Air Travel Clubs as well as many large corporate aircraft certificated under these rules.

Paragraph 121.360 of FAR 121 sets forth the requirements for an acceptable GPWS. In general, the system was to provide for four warning modes:

- Mode 1 - Excessive Sink Rate Close To Terrain
- Mode 2 - Excessive Closure Rate Close To Terrain
- Mode 3 - Negative Climb After Takeoff
- Mode 4 - Descending Into Terrain With Gear Or Flaps Up

The acceptable GPWS performance was further outlined in an Advisory Circular issued on December 31, 1974 (FAA, 1974b).

Numerous additional amendments affected the requirements for GPWS design and use by: (1) defining specific technical performance and environmental standards for GPWS systems (FAA, 1975a); (2) incorporating Mode 5, Operation Below the Glide Slope, by June 1, 1976; (3) extending some GPWS installation dates until June 1, 1976, in recognition of delays in equipment deliveries; (4) waiving a provision prohibiting deactivation of the GPWS system, until September 1, 1976, in recognition of equipment reliability problems and attendant false alarms; (5) redefining the aircraft covered by the ruling, exempting the required GPWS system for aircraft with up to 30 passenger seats, a maximum payload up to 7,500 pounds, and a maximum zero-fuel weight up to 35,000 pounds; and (6) further revising equipment standards, involving alteration of the warning envelopes, to reduce false and nuisance warnings (FAA, 1976).

In 1976/1977 a separate, ground-based warning system (also intended to prevent CFIT accidents) was implemented. This system, the MSAW element of ARTS-3, had been under development for the FAA since the early 1970s. MSAW became operational at Los Angeles and Washington, D.C., in November 1976; initial operation at Denver, Detroit, Houston, and St. Louis was in January 1977, and at the remaining 57 ARTS-3 terminals by mid 1977.

With MSAW the air traffic controller receives a visual signal and an aural alarm when an eligible aircraft penetrates or is predicted to penetrate (by extrapolating descent rates) a predetermined minimum safe altitude in the terminal area. MSAW operates in two modes. The one provides for surveillance in all parts of the terminal area except the final approach corridor. The second is tailored to monitor aircraft altitude versus position during the final approach.

ISSUES AND ATTITUDES

The initial GPWS ruling was issued just days after the crash of Trans World Flight 514--a Boeing 727--into a mountain near Dulles International Airport. This was but one in a succession of CFIT accidents in the late 1960s and early 1970s. The FAA had been evaluating airborne warning systems relating to CFIT accidents and was planning to adopt a GPWS ruling in April 1975. However, the pressures following the Flight 514 accident prompted the FAA to act in December 1974.

Prior to 1975, accidents in which airliners flew into the ground while in controlled flight had been a leading cause of accidents. According to McDermott (1975), over 50 percent of all airline crashes in the prior 20 years had fallen into that category. Bateman found 115 CFIT-type airline accidents for use as case studies (Bateman, 1972). In part, radar altimeters and altitude alert systems (barometric altimeter) were developed because of a concern for such accidents. However, efforts to develop a comprehensive GPSW didn't start until 1966.

Advocacy for a mandatory domestic GPWS system probably began with the NTSB about 1970. Subsequently, the Airline Pilots Association (ALPA) advocated such systems. By 1974, the Boeing Airplane Company was planning to install GPWS in new aircraft. In May of 1974, following its fourth fatal CFIT accident in three years, Pan American World Airways initiated a program to equip its entire fleet with GPWS systems. In September 1974, prior to the TWA 514 accident, FAA had announced intent to adopt a GPWS ruling in early 1975.

There were legitimate doubts about the need for, or the prospective effectiveness of, such systems. Some were skeptical of GPWS effectiveness in cockpits already containing 200 alert and warning lights and 20 or more warning bells, buzzers, and sirens (McDermott, 1975). They speculated that if it was effective, it would be because unlike the other warning systems it (1) couldn't be ignored and (2) couldn't be turned off. The FAA's skepticism was clear in its original amendment requiring GPWS (FAA, 1974a):

> "The FAA believes that present instrumentation and inflight procedures provide for safe and adequate terrain clearance as long as proper flight crew member discipline is maintained and appropriate flight operations procedures are followed. However, notwithstanding those instruments and procedures . . . a number of air carrier accidents involving large turbine-powered airplanes have been caused by inadvertent contact with the ground, and might have been avoided if a ground proximity warning system had been

> installed to give warning of the impending disaster to the flight crew."

Nor were these sectors alone in their skepticism. In a technical paper presented in early 1977, one human-factors expert wrote:

> "An appealing, though perhaps somewhat simplistic, answer to vigilance and attention problems is to install warning devices which are assumed to be attention demanding . . . There is never a guarantee that any stimulus, regardless of its psychophysical dimensions, will be correctly attended to 'correctly' because one of the usual ways of dealing with warning devices, particularly in the auditory mode, is to shut them off. The proposed ground proximity warning systems in the cockpit may be well intentioned, and may indeed prevent some CFIT accidents, but human factors specialists would do well to regard them with a measure of skepticism. (Wiener, 1977)"

GPWS/MSAW IMPACT ON CFIT ACCIDENTS

To our knowledge, the impact of GPWS and MSAW implementation has not been examined heretofore. That does not deny a general awareness in the aviation community of fewer CFIT accidents. Bateman (1978) noted the reduced airline CFIT accident fatalities during the early years following GPWS introduction and the continuing fatal accidents experienced by unequipped corporate fleets and U.S. military transports.

To quantify the impact of GPWS and MSAW systems on CFIT accidents, we examined NTSB accident data for 1971 through 1980, including unpublished data for 1979 and 1980 from internal NTSB records.

To qualify as an "appropriate" CFIT accident, an accident had to involve the aircraft and operations types affected by the GPWS regulations.* That is, it must involve a large turbine-powered aircraft operated under FARs 121, 123, or 135. Accident fatalities were not a criterion; the interest is in accidents of the CFIT type, not their survivability. In fact, major airframe damage was not even a criterion. Thus, as long as the aircraft made contact with ground, water, or obstacles thereon, the accident was of interest--even if the crew was able to get the aircraft airborne again following impact.

The more subtle criterion was that prior to impact the events in the cockpit qualify the accident as Controlled-Flight-Into-Terrain defined as:

> "Controlled-flight-into-terrain accidents are those in which an aircraft, under the control of the crew, is flown into terrain (or

---

*This approach might fail to capture aircraft not required to have GPWS but equipped with altitude-reporting transponders and whose accident frequency might be affected by MSAW alone.

water or obstacles) with no prior awareness on the part of the crew of the impending disaster."

Many accidents in the NTSB files involving hard landings, and runway undershoots were excluded because the evidence of poor piloting technique or control problems brought about by other causes disqualified them from being considered "controlled flight".

In the 10-year period of interest, 19 accidents were identified which met the criteria discussed above (See Table 1). It was observed that the number of accidents in the years following GPWS and MSAW implementation were markedly reduced as shown in Figure 1. If the study period is divided--1971 through 1975, and 1976 through 1980--the second period can be reasonably labeled as post-GPWS. By December 1, 1975, it is estimated that approximately 80 percent of the required installations were completed. While the ensuing period was marked by some deactivations due to false alarms, most systems were installed and functioning by September 1, 1976. Because MSAW implementation was from November 1976 through June 1977, it is somewhat less precise to call the second period a post-MSAW period.

Seventeen qualifiable CFIT accidents occurred during the first period. By contrast, only two accidents occurred in the second period. This is a statistically significant reduction. By one test*, the probability of this being a coincidental reduction is less than seven in one million. This dramatic reduction cannot be attributed to decreases in aviation activity. In fact, except for a 7 percent drop from 1973 to 1974, activity increased in every year from 1971 through 1979 (FAA, 1979). Thus, the authors see no reasonable alternative to concluding that these two warning systems have had a profound effect in reducing CFIT-type accidents.

The two CFIT accidents during the second period occurred in 1977 and 1978. Both aircraft were equipped with GPWS systems. The 1977 accident involved a United Airlines DC-8F cargo flight near Kaysville, Utah. The 1978 accident was the more publicized crash of a National Airlines B-727 in Escambia Bay near Pensacola, Florida. In the United accident, the GPWS is believed to have given an alert prior to impact--an inoperative Cockpit Voice Recorder precluded determining this with certainty. In the latter case, the GPWS did provide an alert. The alert was ignored. The problem was compounded by the Flight Engineer's action of shutting the GPWS off following the alarm without the pilot's knowledge.

---

*Assuming that the number of accidents in any 5-year period is Poisson distributed, then in the period 1971-1975, the number of accidents is Poisson distributed with a mean rate of 17 accidents per 5-year period. Given that rate, the probability of observing two or less accidents in the second 5-year period is calculated as being 6.7×10^{-6}.

TABLE 1
Identified CFIT Accidents During the Period 1971 Through 1980 From NTSB Data

Date and Location	Description	NTSB Report No.
02-17-71 Gulfport, MS	Southern Airways DC-9-15 struck electrical transmission line 8,900' from runway during VOR Approach to Runway 13	AAR-71-14
06-07-71 New Haven, C[illegible]	Allegheny Airlines Convair 580 struck beach cottages 4,890' from runway during a non-precision approach to Runway 2	AAR-72-20
06-22-71 Martha's Vineyard, MA	Northwest Airlines DC-9-31 struck water 3 miles short of runway during VOR Approach to Runway 24	AAR-72-4
07-25-71 Manila, P.I.	Pan American Airways Boeing 707 hit mountain far from airport following premature descent in preparation for ILS Approach	Unpub.
09-04-71 Juneau, AK	Alaska Airlines Boeing 727 struck mountain 18.5 miles from airport during Localizer Approach to Runway 8	AAR-72-28
12-29-72 Miami, FL	Eastern Airlines L-1011 struck ground during unintentional descent while holding at 2,000' 18 miles from airport	AAR-73-14
04-10-73 Toledo, OH	Eastern Airlines Boeing 727 struck trees 6,400' short of runway during a Localizer Backcourse Approach to Runway 25	AAR-73-17
07-22-73 Papeete, Tahiti	Pan American Airways Boeing 707 struck water 3 miles from airport while in a departure turn from Runway 4	Unpub.
07-31-73 Boston, MA	Delta Airlines DC-9 struck seawall 3,000' from runway during ILS Approach to Runway 4R	AAR-74-3
09-08-73 King Cove, AK	World Airways DC-8 struck mountain 15.5 miles from airport following a descent preparatory to making an instrument approach	AAR-74-6
09-27-73 Mena, AR	Texas Int'l Convair 600 struck mountain during cruising flight	AAR-74-4
01-30-74 Pago Pago, Samoa	Pan American Airways Boeing 707 struck trees 3,865' short of runway during ILS Approach to Runway 5	AAR-74-15 & AAR-77-7
04-22-74 Bali, Indonesia	Pan American Airways Boeing 707 crashed into high ground 37 miles from airport during NDB Approach procedural turn	Unpub.
09-11-74 Charlotte, NC	Eastern Airlines DC-9-31 crashed 3.3 miles from runway during VOR/DME Approach to Runway 36	AAR-75-9
12-01-74 Berryville, VA	Trans World Airlines Boeing 727 struck ridge 25 miles from airport following clearance to VOR/DME Approach to Runway 12	AAR-75-16
08-30-75 Gambell, AK	Wien Air Alaska F-27B struck mountain while circling to land following a missed approach	AAR-76-1
11-12-75 Raleigh, NC	Eastern Airlines Boeing 727 struck ground 282' short of runway during an ILS Approach to Runway 23	AAR-76-15
12-18-77 Kaysville, UT	United Airlines DC-8F crashed into mountain while holding in preparation to landing	AAR-78-8
05-08-78 Pensacola, FL	National Airlines Boeing 727 crashed into water 3 miles from runway during Radar Approach to Runway 25	AAR-78-13

GPWS/MSAW EFFECT ON THE OPERATIONAL DOMAIN

Traditionally, the ability to acquire significant quantities of useful information from the operational aviation world--especially that pertaining to human error--has been quite limited. Bateman (1978), in his paper assessing early performance of GPWS systems, says:

> "What about reported saves? It would be the rare individual pilot, indeed, who would ever inform his chief pilot or company president that he almost flew the company's airplane into a mountain top, save for a timely, GPWS warning (let alone a copilot or flight engineer's shout of 'Pull Up'). This is human nature. On the other hand, we will surely hear of any nuisance alarm or false warning.

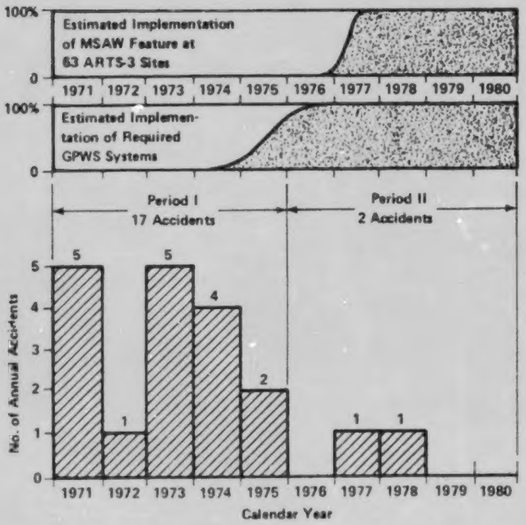


FIGURE 1. Frequency of Controlled-Flight-Into-Terrain Accidents Before and After GPWS/MSAW Implementation

> As a result of this human element, the reported incidents of GPWS warning will normally be those where the aircraft position or unsafe flight path is caused by factors other than crew judgement. An example is improper ATC instruction."

This situation was dramatically altered with the advent of the ASRS in 1976. This voluntary reporting system offers both confidentiality and limited immunity to the reporter in exchange for information on discrepancies and deficiencies in the national aviation system. A large fraction of the 24,000-plus reports received to date are self-incriminating to the reporter.

A recent ASRS study "An Investigation of Reports of Controlled Flight Toward Terrain (CFTT)" focused on reports from July 1976 through October 1980. For the period, the ASRS data base contains about 23,000 reports. Of these, 266 reports outline occurrences or situations providing insight into the CFTT problem area and 51 specifically mention either GPWS or MSAW. While the 266 reports involve general aviation and scheduled air carriers, the 51 GPWS/MSAW reports deal almost exclusively with air carrier operations.* Some imply that the GPWS/MSAW alarm was appropriate; others allege the resulting alarm to have been false or a nuisance (see Table 2).

---

*The 22 MSAW alarms contained one general aviation and one military case.

The 22 valid GPWS or MSAW alarms include six where the GPWS or MSAW gave the first warning of trouble and where, due to other factors, the authors believe a CFIT accident would have otherwise resulted. GPWS and MSAW each accounted for three of these probable saves. Two other incidents, reported after the CFIT study was completed, fit in this category. Both involved MSAW alerts. One is an ASRS report received in November 1980; the other is the well-publicized incident involving the Argentine Airlines Boeing 707 on a collision course with the 110-story World Trade Center in New York City on February 20, 1981. The two incidents described below indicate the dramatic nature of some of these probable saves. The first example, a "GPWS save" in February 1977, at Hill AFB, Utah, involved an air carrier transport. The incident occurred during daylight hours and under Instrument Meteorological Conditions (IMC).

> "The GPWS actuated with a red light and 'whoop-whoop pull-up'. At this time, we were on a radar vector by HIF Approach at an assigned altitude of 11,000 feet. Vector heading assigned at that time was 250 degrees. Radio altimeter was observed to pass 2,500 feet rapidly, and power was initiated, climb attitude established. The radio altimeter passed through 800 feet and gradually started up during the climb. After passing the crest of the mountain, everything returned to normal."

TABLE 2
Distribution of Reported GPWS/MSAW Alarms

System	Total Reports	Validity of Alarm	
		Valid	False/Nuisance
GPWS	28	12	16
MSAW	22	10	12
BOTH	1	0	1
	51	22	29

The assigned 11,000 foot altitude was the normal Minimum Vectoring Altitude (MVA) in use. Ordinarily, it would have provided adequate clearance of a 9,674 foot MSL mountain peak in route. However, following landing, an investigation revealed a deep low pressure trough aloft. Because of it, and its effect on the barometric altimeter, the aircraft was probably 1,400 feet lower than its altimeters indicated. The MVA for that area was subsequently raised to 12,000 feet.

The second example, an "MSAW save" in November 1980, near Tampa, Florida, involved a large air carrier transport. The incident occurred at night under IMC conditions.

> "The aircraft was executing a Localizer Back Course Approach to Runway 18R at Tampa International Airport. Keno, which is the final approach fix on the approach, is located 5.5 miles from the airport and has a mandatory crossing altitude of 2,000 feet. The aircraft made his initial call up to Tampa Tower stating that he was coming up on Keno. At this time, he was cleared to land, which he then acknowledged and asked if the visibility was still holding at a couple of miles. He was told the visibility was 5 miles and

> he acknowledged. I received a low altitude alarm for him as he was about 4 miles from the airport and his altitude indicated 300 feet. I immediately issued a low altitude alert advising him of his position and his altitude which now indicated 200 feet. He acknowledged and said he was checking it."

In a November 19, 1975 letter, the Air Transport Association of America (ATA) petitioned the FAA to amend FAR paragraphs 121.303 and 121.360 to provide relief until September 1, 1976, from the requirement that the ground proximity warning system required by paragraph 121.360 be in operable condition before takeoff. This petition was motivated by false alarm problems experienced by ATA's member airlines. In its amendment acting on this petition FAA quoted parts of the ATA letter (FAA, 1975b).

> "At least two major carriers are experiencing an unacceptably high number of false and nuisance alarms. More specifically, a false warning occurs when the system alarms, although system parameters do not call for one. For example, warnings have occurred on takeoff when the alarm sounds while the aircraft is in a positive rate of climb well outside the TSO envelopes which would require the warning. Warnings have occurred in cruise at high altitude for unexplained reasons, as well as during approach when the aircraft is on the localizer and glide slope over flat terrain."

The ATA was concerned with the effect of numerous false and nuisance warnings.

> "Pilots will quickly lose confidence in this system if this continues for even a short period of time. Once they lose confidence, it will be practically impossible to regain. Then, the efforts of both FAA and industry to realize the safety benefits which this system promises will have gone for nothing. We will have spent thousands of man-hours and millions of dollars on a black box that nobody trusts."

As reported earlier, the ASRS system received reports detailing both GPWS and MSAW false or nuisance alarms. While their number is not a valid statistical indicator of the actual number of false alarms experienced, the reports provide some useful insight into the phenomenon (see Table 3).

While false alarms for GPWS and MSAW occurred in all phases of flight, most occurred in the final approach phase (see Table 4).

Of the 29 incidents, 12 occurred in visual meteorological conditions (VMC), 8 in IMC. The remainder were not identified. The amount of VMC operation--in contrast with that in IMC--may influence the number of reports received in that category. The high number of VMC incidents may also result from the wider range of flight paths flown in VMC, as opposed to precision instrument approach flight paths. The probability of violating the alarm envelopes of these systems would be greater under such circumstances.

TABLE 3
Distribution of Reported False or Nuisance GPWS/ MSAW Alarms by System and Year

Calendar Year	Total Reports	Warning System		
		GPWS	MSAW	Both
1976*	5	5	0	0
1977	6	2	4	0
1978	11	6	4	1
1979	4	2	2	0
1980**	3	1	2	0
	29	16	12	1

*includes only the last 6 months

**includes only the first 10 months

TABLE 4
Distribution of Reported False or Nuisance GPWS/ MSAW Alarms by System and Phase of Flight

Phase of Flight	Total Reports	Warning System		
		GPWS	MSAW	Both
Climb	1	1	0	0
Cruise	1	1	0	0
Descent	3	2	0	1
Initial Approach	1	1	0	0
Final Approach	22	10	12	0
Unknown	1	1	0	0
	29	16	12	1

The reporters of these incidents--83 percent pilots--expressed various concerns about false or nuisance alarms. These concerns indicated four hazards: (1) encountering a midair collision while performing a mandatory pull-up, (2) losing control of the aircraft while distracted, (3) ignoring a valid alarm because of system credibility problems, and (4) ignoring a valid alarm through a misunderstanding of exactly what triggered it.

In four reported cases the flight responded to a GPWS alarm with the required pull-up maneuver. Each alarm was subsequently discovered to have been false. Since each such uncoordinated pull-up has some risk of a midair collision, the pilots wonder if the cure isn't worse than the disease.

In many cases the distraction of the alarms--both MSAW and GPWS--was cited as the greatest concern. That such distraction could lead to a loss of control was realized in two reported cases. In a daylight visual approach to San Francisco International Runway 28L, full flap extension could not be achieved in the late stages of the final approach. Because of other traffic and a fog bank in the missed-approach corridor, a decision was made to continue the approach to landing.

> "Increasing the sink rate to approximately 1,500 feet per minute and spooling the engines simultaneously provides a good stable platform two miles from the end of the runway. This was the maneuver that was in progress at the time the ground proximity warning system was triggered to a pull-up alert because of the increased sink rate. Since there is no means of silencing the GPWS system and since the audio level is extremely high, all cockpit and tower communications were blocked at this point in time. Speed callouts, sink rate callouts, height above the runway callouts were not possible. This, in effect, denied a coordinated crew function. The landing was effected without incident."

One terse pilot report tells the story of a daylight VMC approach to Honolulu.

> "GPWS malfunction in gusty wind conditions on short final, on glide slope, gear down, full flaps (50 degress), speed ^{V}Ref + 10 knots, GPWS says 'too low, flaps'. A quick scan showed everything OK. Alert second officer pulled circuit breaker. Distraction caused airspeed to go below ^{V}Ref at a time when recovery was difficult. False signals from GPWS, which occur often, create such a distraction that flight safety is impaired."

Numerous reporters expressed concern that false alarms hurt the system's credibility and lead to ignoring it when valid. Two dramatic cases of ignored valid alarms demonstrate this type of problem. The first is a night approach to Tampa International Airport under VMC conditions.

> "Approach to 18R using VASI was stabilized and in visual glide path. At approximately 300 to 400 feet MSL, ground proximity warning system activated giving a 'pull-up'. Immediate reaction was 'what the hell is causing it to go off this time', meaning in my mind 'what's wrong with the GPWS', not 'what is it trying to tell me'. I have experienced literally dozens of these warnings where there was no significance to the warning. I am conditioned. I did glance over to see that the landing gear was indicating down and locked and the First Officer pushed the glide path cancel light. But we continued to touch down. During roll out I discovered that the flaps were at 25 degrees, which is not a landing flap . . ."

The second example involves a nighttime approach to Kansas City Airport under VMC conditions. After planning for an approach to Runway 01, the flight was granted permission for a more expedient approach to Runway 19. The clearance came late, and the aircraft had to be maneuvered significantly to make the change.

> "I was prepared for a steep initial descent in a short turn-in, had frequency and localizer course at hand and already selected. I kept the Flight Director bars centered through the approach. The GPWS alarm went off but was ignored as giving erroneous warnings . . . normal landing followed. Not until later did I realize I had the wrong ILS frequency selected."

Alarms have been construed as false because the flight crew did not understand which factor triggered the alert. This has happened in MSAW alerts where the tower has advised of a "low altitude" alert. The flight crew checks the altitude only to find that it is normal when excessive rate of sink was the cause of the alarm.

CONCLUSIONS

GPWS and MSAW warning systems have brought about a dramatic reduction in CFIT accidents and undoubtedly have saved hundreds of lives. GPWS or MSAW was likely the sole factor in the prevention of several accidents; in other cases, while accompanied or followed by other warnings, their role was likely decisive in preventing the accident.

Although false alarms are a nuisance and cause for concern, there is no evidence of any accident having been caused by such false alarm.

Many pilots and controllers are not aware of the net benefit of the GPWS/MSAW systems. One ASRS report reflects a view held by many: "In my opinion and also the opinion of my fellow pilots, the GPWS is more likely to cause an accident than to prevent one."

The efficacy of the GPWS system, versus the perceived value of warning devices held by some, probably traces to its two unique characteristics - it can't be ignored and it can't be disarmed with anonymity. How much room this leaves for adding other tenacious alarms remains to be seen.

In the face of serious hazards to flight (e.g., the rash of CFIT accidents prior to 1975), it may not be optimal or desirable to carry prevention device or procedures development to perfection before implementation.

The valid (alarm criteria are met) but nuisance alarms frequently benefit flight safety by inducing the crew to examine, in retrospect, the conditions leading to the alarm. This has led to the detection of system problems or deficiencies.

REFERENCES

Bateman, C. Don, 1972, Effectivity Analysis - Ground Proximity Warning, Sundstrand Data Control, Inc., Redmund, Washington, Report 070-0723-001, September 18.

Bateman, C. Don, 1978, "Effect of GPWS on Terminal Area Accidents", Proceedings of the 23rd Corporate Aviation Safety Seminar, Flight Safety Foundation, Inc., Arlington, Virginia, April 9-12.

(FAA) Federal Aviation Administration, 1974a, "Ground Proximity Warning Systems", Federal Aviation Regulations, Part 121, Amendment 121-114, Government Printing Office, Washington, D.C., December 18.

Federal Aviation Administration, 1974b, Ground Proximity Warning Systems, FAA Advisory Circular No. 25-6, Government Printing Office, Washington, D.C., December 31.

Federal Aviation Administration, 1975a, "Technical Standard Order for Ground Proximity Warning Equipment (TSO-C92)", Federal Aviation Regulations, Part 37, Government Printing Office, Washington, D.C., September 11.

Federal Aviation Administration, 1975b, "Ground Proximity Warning Systems", Federal Aviation Regulations, Part 121, Amendment 121-126, Government Printing Office, Washington, D.C., November 24.

Federal Aviation Administration, 1976, "Ground Proximity Warning -Glide Slope Deviation Alerting System", Federal Aviation Regulations, Part 121, Amendment 121-129, Government Printing Office, Washington, D.C., August 13.

Federal Aviation Administration, 1979, *FAA Statistical Handbook of Aviation, Calendar Year 1979*, Government Printing Office, Washington, D.C.

McDermott, Jim, 1975, "FAA Pushing New Electronic Aids to Prevent Flight Into the Ground", *Electronic Design 11*, May 24.

Wiener, Earl L., 1977, "Controlled Flight Into Terrain Accidents: System-Induced Errors", *Human Factors*, Volume 19(2), April, pp 171-181.

ASSESSING EMERGENCY INTERFACE DESIGN

William C. Allen
Stanford University

INTRODUCTION

A functional model of beam dynamics has a theoretical basis in elasticity and mechanics. Similarly, a functional model of man/machine interaction must have an underlying theoretical basis for how the man and the machine respond to their environments. In most circumstances, control theory provides an adequate theoretical model for the mechanical system. Efforts to adapt this theory to the human operator have encountered severe limitations -- especially in unfamiliar or multi-attribute control environments (i.e., emergencies or strategic level decision making).

If we designed buildings using an elasticity model valid over a very narrow region, we would be careful not to employ this model as a predictor of system response outside that region (otherwise our building might fall down). We can criticize almost all existing interface design procedures because they assume (usually implicitly) that an optimal control model for operator response is valid over the entire range of control environments. Therefore, it should not be surprising that performance of these systems degrades seriously when exposed to situations where this assumption is invalid.

An important byproduct of this realization is that if an explicit model of human response is not incorporated into a design procedure, an implicit model will take its place. Clearly, something as critical as operator response should not be left to default modelling.

Cognitive psychologists have developed a general theory for human information perception, storage, retrieval and manipulation based on the notion of schemata. One difficulty with modelling human response using schema theory is that there is little understanding of the underlying mechanisms of schema manipulation. However, in the same way that it is not necessary to understand the nuances of molecular interactions to make predictive theories about materials behavior, it can be hoped that we can address human response tendancies based on a macro level schema theory.

The next section highlights the fundamentals of schema theory necessary for this discussion. The third section addresses how the process of schema selection optimization appears to represent a reasonable descriptive model of operator response. In the fourth section, this model is expanded to provide an analytical approach to assessing interface design.

The result is a structured approach to addressing human decision making/action selection over a wide range of operating environments. Simulation can be used to illuminate the different paths the inferential decision maker can take given a particular display/control format and a given set of status/warning indications.

It is hoped that future work in this area will allow this modelling/simulation/analysis process to be codified. The result would be a general design procedure for emergency interfaces.

Introduction to Schemata

By constraining system design to include a "man in the loop", we are forced to recognize and address the issues of non-normative decision making typical of human response. Schema theory provides a robust mechanism through which we can address human information processing. Three structures must exist for schema theory to function:

1. Semantic knowledge structures. These are semantic networks between key variables in a frequently encountered phenomenon.

2. Episodic memory structures. The key features in an experience (as well as connectors between the features and the moderating schemata) are retained as episodic memory traces.

3. Schema selection optimization. Two levels of optimization must occur:

    - -- The relative costs of delays and potential perception errors must be rapidly balanced when selecting the schema to be used to structure exogenous data.
    - -- The costs of validating and determining the uniqueness of an "apparently valid" schema must be optimized before it is used to generate response scenarios.

Semantic and episodic knowledge structures are highly interactive and mutually supportive. Semantic memory provides the framework used to encode and recall episodic memory traces. Episodic memory traces provide experiential richness to a partially instantiated schema by allowing experience with "similar" or analogous situations to be brought into the evaluation.

With this approach, we would predict that an operator would deal with limited sensory data by reaching into his episodic memory stores for "representative" values and constraints for unavailable variables. This ability to draw from past experiences provides added "context" when evaluating a situation. Conversely, faded episodic memory traces can be "reconstructed" using the implied contexts and constraints available through the semantic knowledge structure.

Optimizing Schema Selection

Limitations in short term memory dictate that people do not simply load all available sensory data into a memory buffer and perform an exhaustive sort of schemata to determine the best fit. Instead, we expect that a "reasonably valid" schema is rapidly chosen and used as a framework to organize the sensory data. If a selected schema fails to provide an adequate fit for subsequent sensory data, it is rejected in favor of a new schema.

From this, we would expect that inhibition of invalid schemata must play an important role in the initial schema selection process. Otherwise, we would be bombarded with potentially invalid schemata and have to conciously sort through them all. Therefore, the attended features which appear central (and their apparent semantic connections) must serve to inhibit all schemata whose variables or structures do not permit such features.

Since this sort takes place very rapidly, a hierarchy of the variables in a schema must exist. This initial sort must compare the "central" sensory features to schemata with similar central variable constraints.

Based on these (and other) considerations, a "branch and bound" optimization strategy appears to be an appropriate model for the initial schema selection optimization process. This approach allows schema selection to be dependent on both previously attended information and prior decision path. Experiential and analytical heuristics are used to estimate the schema which "appears" to contain the greatest potential for providing a solution. Other heuristics are used to estimate the uniqueness of the solutions resulting from the use of that schema.

Before modifying the branch and bound approach to accommodate schema optimization, it is worthwhile to review the steps in the general algorithm:

1. Partition the solution place into mutually exclusive, collectively exhaustive sub-spaces.

2. Develop upper bounds (assuming a maximization problem) for each of the sub-spaces.

3. Test each upper bound solution to determine if it is feasible.

4. If one (or more) feasible solution exists, make the highest of these the incumbent solution.

5. If any of the following are true, the sub-space is considered fathomed:

    -- subspaces whose upper bound solution is also feasible
    -- sub-spaces with upper bounds lower than the incumbent solution
    -- sub-spaces containing no feasible solutions

    Fathomed sub-spaces can be removed from further consideration.

6. Take either the highest (the most promising) or the most recent of the unfathomed sub-spaces (depending on the branching rule employed) and partition it into smaller sub-spaces.

7. Go back to step 2 and continue until all subspaces are fathomed -- the incumbent solution at this point will be an optimal solution.

The schema selection algorithm would follow the same general pattern:

1. The "solution space" is partitioned by activating schemata with appropriately constrained "central" variables and inhibiting inappropriate schemata. Within these overlapping sub-spaces simple cause-effect relationships exist between focal and peripheral components or sub-systems.

2. Constraints provided by episodic memory are used to provide "quick access" bounds on the "goodness of fit" of all active schemata.

3. The best "quick fit" would be selected as a candidate schema. Using available features to partially instantiate the candidate would form a test of it's feasibility (i.e., if structural or variable constraints were violated, the schema would be fathomed and removed from consideration.

4. The "stopping rule" for the schema search algorithm would be based on either the difference between the "goodness of fit" of the partially instantiated candidate schema and other active schemata or the difference between the episodic associations activated by the partially instantiated candidate and those activated by the available features (i.e., improved "goodness of fit" of the partially instantiated schema inhibits activation of alternate schemata while poor "goodness of fit" increases their activation).

When applied to schema selection, branching difficulty can be considered the number of hypotheses which are "candidates" for interpreting the data. If only one hypothesis receives activation from the available (or attended) features, little or no cognitive effort is required. These unambiguous hypothesis selections should appear "automatic" (i.e., require no cognitive effort).

Conversely, if multiple hypotheses (or no hypothesis) receive activation, conscious hypothesis selection must occur since *some* means must exist for organizing the available data. For example, if conscious hypothesis selection does not take place in an ambiguous situation, the operator will be unable to make "sense" of it and will take no action.

This discussion implies that ambiguous situations will be predominated by "conceptually driven processing" while unambiguous situations will be predominated by "data driven processing". This points out a significant difference between designing routine and emergency interfaces:

-- Routine situations can be effectively controlled by providing adequate sensory input (data driven processing predominates).
-- Emergency situations require simulation of appropriate perceptions of the situation to allow adequate response (conceptually driven processing predominates).

Since initial schema selection is used to encode available data, while action selection involves an irrevocable allocation of resources, we would expect additional schema optimization testing before action is allowed. Although the incumbent schema can be considered reasonable and feasible after it's acceptance as a sensory data framework, it's uniqueness remains to be resolved.

In unambiguous situations, this fathoming of alternate candidates is trival -- there are no alternate candidates. Therefore, we can expect that these situations can lead to "slips" when relatively rare alternatives are the "correct" choice. For example, if an indicator "always" implies a particular action during normal system operation, we can expect "intrusions" of that action during emergency operations where the implications of that indicator may be more ambiguous.

This approach asserts that the schema will remain unchanged unless contradictory evidence is found. Since evidence is asserted to be requested for the purpose of confirming the current hypothesis, we can expect that an incorrect hypothesis formed on the basis of weak evidence will be more resistant to change based on new, better data than a hypothesis formed on the basis of the new data alone. For example, when critical data is unavailable, several schemata might be selected with equal validity. Exposure to a few non-critical features contradictory to the intial schema selection will likely result in the operator "explaining away" the new features rather than changing schemata.

Assessing Interface Design

Although schema theory provides valuable insights, it lacks the computational structure necessary for quantitative analysis. The branch and bound model for schema manipulation provides the needed structure but does not include a systematic means for handling uncertainty in our knowledge of the interactions between model parameters or uncertain outcomes.

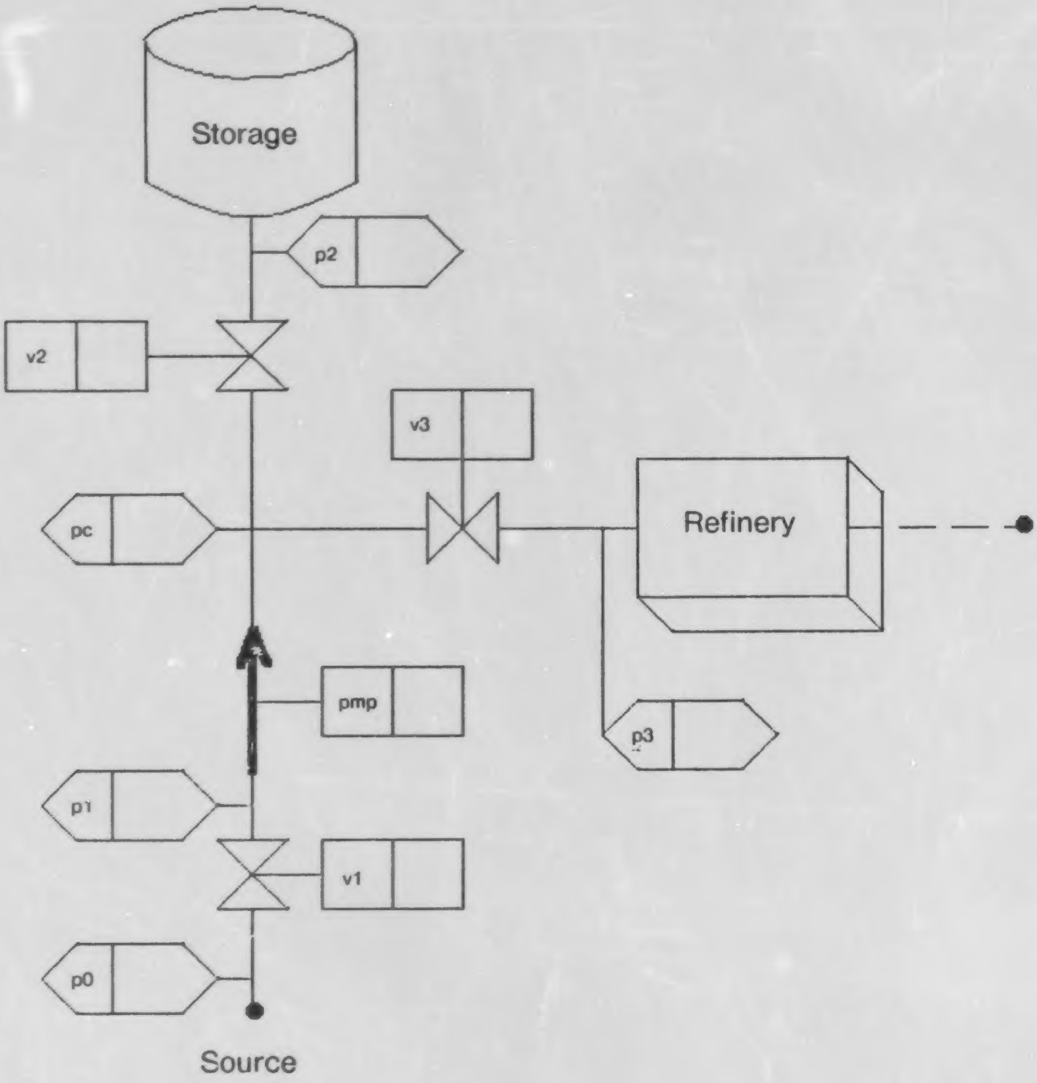
This shortcoming can be resolved (to some extent) by adopting a modified decision analysis approach to interface design assessment. Since the structural models used in decision analysis share with branch and bound a sequential tree-like structure, the models should be compatible.

Figure 1 depicts a "status annunciating" display configuration for a simple oil transport/storage system. This system assumes that the oil source (i.e., tankers) is to be unloaded as quickly as possible and that excess pumping capacity is diverted to temporary oil storage tanks. The valves can be open or closed, pressures can be zero, low, medium or high and the pump can be set at off, slow or fast. If problems develop, they will be annunciated by the lights at the bottom of the display.

Figure 2 is the abbreviated decision tree which represents a simplified structural model for an inferential decision maker. Note that the structural model follows the same general format outlined in the discussion of schema theory but excludes all functional details. Figure 3 is a truncated decision tree which results when this model is applied to the example display format.

The dependence of decision variables on attended information and the heuristic nature of the hypothesis formation process causes the "decision" variables to retain a probabilistic nature. The probability associated with the selection of a candidate "decision" depends on the "degree of association" existing between the candidate and the hypothesis given a particular generating scenario. This approach is consistent with the activation model for schema moderated behavior presented in the previous section.

At this level of sophistication in the inferential decision model, the assignment of probabilities to the branches of a decision node would be entirely subjective. "Reasonable" values for these probabilities would be derived from experience with similar systems, results of pencil and paper or simulator studies, or from "engineering judgement".

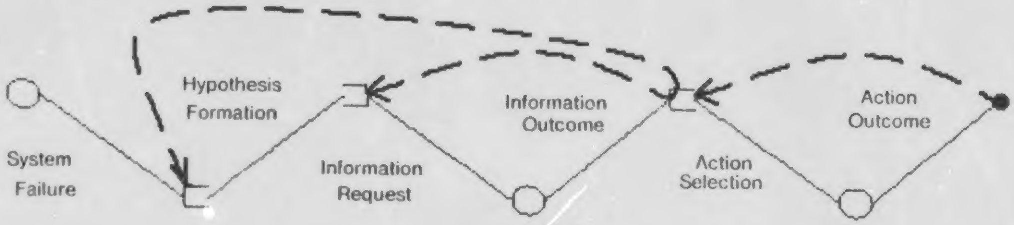

Storage
p2
v2
v3
Refinery
pc
pmp
p3
p1
v1
p0
Source

WARNINGS

Source Disruption	Pump Breaker Tripped	Excessive Tank Pressure	Insufficient Production Pressure

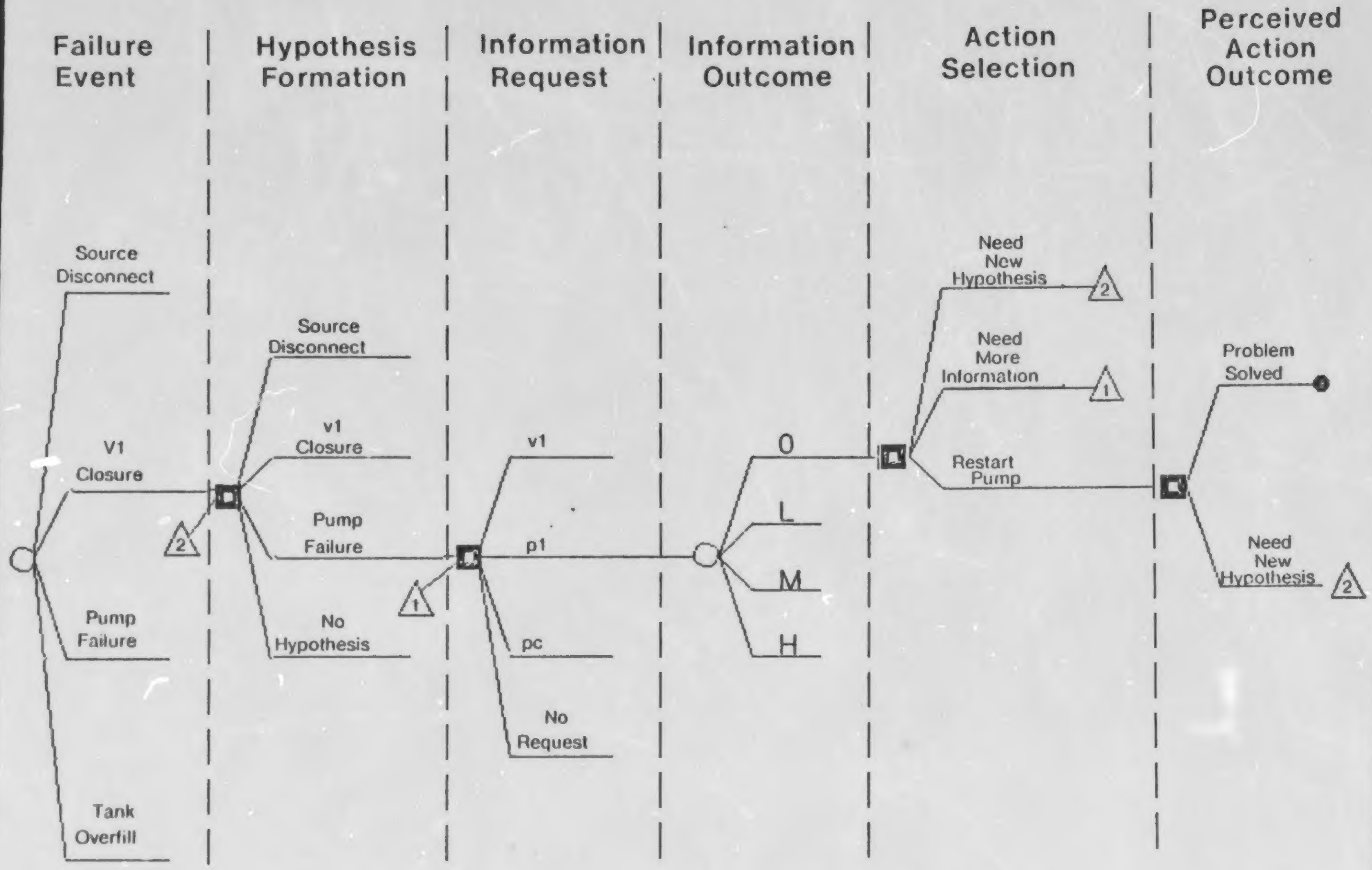
Status Warning Annunciator Display

Figure 1


Hypothesis Formation
Information Outcome
Action Outcome
System Failure
Information Request
Action Selection

Abbreviated Inferential Decision Tree

Figure 2


Failure Event
Hypothesis Formation
Information Request
Information Outcome
Action Selection
Perceived Action Outcome
Source Disconnect
V1 Closure
Pump Failure
Tank Overfill
2
Source Disconnect
v1 Closure
Pump Failure
No Hypothesis
1
v1
p1
pc
No Request
0
L
M
H
Need New Hypothesis
2
Need More Information
1
Restart Pump
Problem Solved
Need New Hypothesis
2

Truncated Inferential Decision Tree
Figure 3

Additional detail can be incorporated into the model by including an "attended features" branch prior to schema selection (as schema theory would dictate). In this way, probability assignments for schema selection would be conditional on the attended features and/or the contents of short-term memory immediately prior to schema selection.

Although this would reduce the subjectivity of the schema selection probability estimates, it greatly increases the computational difficulty of the problem by necessitating the inclusion of a significant number of "attended feature" branches into the model (the number of branches would be approximately the number of observable features chosen the size of working memory at a time -- a huge number for any realistic situation.

To a large extent, this difficulty can be overcome by eliminating trivial or redundant branches. However, a more likely solution would be expected to lie in adopting a simulation approach to enumerating the effects of various attended feature combinations. Regardless, a great deal of analysis would be required to assess the impact of various attended feature combinations on schema selection.

In the simple model presented, action selection and information request probability estimates would be based on experience, intuition or the results of simple experiments. However, an expanded model could include conditional branches to account for the predictable effects of requested information on schema activation and inhibition. Additional conditional branches could be included to account for the effects of differing schema fathoming or stopping rule strategies on information requests and action selection.

For example, time pressure, experience, environmental distractions and motivation can all have a reasonably predictable influence on these parameters. Conditional branches can be included to allow for these variations. As system reliability is allowed to degrade, the dynamic nature of the task is allowed to increase or time sharing activity is increased, additional conditional branches would have to be included to account for the effects.

It is important to make a distinction between computational and theoretical modelling difficulty. The advantage of a structural model is that explicit assumptions are made about the path taken to reach a particular decision or event. The constraints imposed by these preconditions makes it much easier to predict what will happen at that particular point in time. The difficulty lies in the computational burden of going through all the possibilities and estimating what will happen at each.

Therefore, this approach reduces the theoretical complexity of the analysis at the cost of increased computational difficulty. Fortunately, it is far easier to develop computational short-cuts and streamlined algorithm's to solve these computationally more complex problems than it is to develop a comprehensive theory about human information processing.

For example, a progressive model building/sensitivity analysis approach might allow significant "pruning" of redundant or trivial branches before they have to be explicitly evaluated. This would be accomplished by developing a simple structural model, making "ball park" estimates for the possible branches and their probabilities and performing a sensitivity analysis to determine which branches appear to be most important. Less important variables would be set at "nominal values" while the complexity of the model would be increased for the critical variables. Oddly enough, this is exactly the process that schema theory predicts takes place in human information processing.

Simulation can be used to enumerate decision and event chains which present a significant hazard (hazard is defined as the probability of a decision/event chain times the "cost" of the outcome). Additional modelling effort can be given to addressing the interaction of the variables in decision/event chains representing the greatest hazard.

At the very least, this approach has the benefit of explicitly enumerating the man/machine/environment interaction assumptions necessary to design an interface. The combination of formalized structure and simulation aids the designer to uncover unintuitive or insideous sources of operator or system error. Knowledge of biases in human information processing can be used to ferret out potential "slips" or inferential errors which would not be illustrated by an optimal control or sequential analysis approach. Further, the structure of this approach makes it less likely that a designer will assume his own response biases in assessing potential operator response.

The major benefit of an approach of this kind is it's usefulness in overcoming one of the greatest difficulties (and dangers) in a priori caution and warning system design -- in order for problems to be clearly and unambiguously annunciated, the possibility of operator uncertainty at all potential decision points must be foreseen by the designer.

Prospects for the Future

It is entirely possible that cross-situational consistency in certain categories of response can be uncovered by a structured approach such as the one proposed. If these consistencies can be codified into either a simulation or rule based system, it may be possible to evaluate interfaces in a fraction of the time necessary to "start from scratch".

Further, if a system of evaluation can be codified, it can be used to optimize the design of the interface. This presents the possibility of elevating interface design from a "satisfycing" approach to one where the designer can quantitatively estimate how far the proposed system deviates from "optimal". This would allow him to perform "value of information" studies to judge how much additional expense or research effort is warranted in attempting to improve the design.

THE EFFECTS OF ALERT PRIORITIZATION AND INHIBIT LOGIC ON PILOT PERFORMANCE

D. A. Po-Chedley
Douglas Aircraft Company

The purpose of this study was to provide some preliminary insight into the potential benefits that might be associated with alert prioritization and inhibit logic. Prioritization employs an electronic display medium (i.e., a CRT) and sophisticated computer logic to automatically prioritize failure information, thus allowing the flight crew to assess a multiple failure situation in a more expedient manner. Inhibit logic provides the capability to allow the alerting system to inhibit non-essential information during high workload flight segments.

Twelve commercially rated pilots participated, each completing 12 flights during which, they were told, a multiple failure situation might occur. In terms of fault correction time, the use of prioritization resulted in a significant reduction in response times (p .001). The use of inhibit logic also resulted in a noticeable improvement, although not significant across all measurements. Performance on the fault correction task was clearly the poorest when neither prioritization nor inhibit logic was used. Pilots made significantly more response input errors when neither cure was employed. In response to a debriefing questionnaire, the pilots showed a clear preference for the use of prioritization with or without inhibit logic as an aid in identifying the appropriate failure correction sequence. In the absence of prioritization, they exhibited a strong preference for the use of inhibit logic over conventional presentation techniques. Suggestions for future research are made.

7.

PROCRU: A MODEL FOR ANALYZING FLIGHT CREW PROCEDURES IN APPROACH TO LANDING

S. Baron, G. Zacharias, R. Muralidharan, R. Lancraft

Bolt Beranek and Newman Inc.,
10 Moulton Street
Cambridge, Maine 02238

PROCRU, a new simulation model for analyzing crew procedures in approach to landing, is described. The model is a system model that can account for vehicle dynamics, environmental disturbances and crew activities in information processing, decision making, control and communication. Crew sub-tasks are defined based on a time-line analysis of nominal procedures. Information processing and control behavior is modeled after the approach utilized in the Optimal Control Model. Decision making behavior is based on maximizing subjective expected gain. The result is a complex, stochastic model for analyzing the impact on approach and landing of system, procedure and crew variables.

ERGONOMIC ASPECTS IN COCKPIT LAY-OUT

H. AUBAUER W. SPERR

Austrian Flight Safety Board
Vienna, Austria

INTRODUCTION

There is no question that the pilot's role has changed. The days of the autocratic Captain are long past and it seemed more appropriate to call the present era that of the system manager. In this time the importance of the ergonomic fundamentals of cockpit panel design experienced a further increase.

Today there is much more to ergonomics in a man-machine system design than meets the eye. Yet the displays, controls and panel design, like the tip of an iceberg to a seaman, are almost all the pilot is consciously aware of in the application of human factors engineering to his flying task.

Although panel design is only a part of the ergonomics package it is one that should not be neglected. All those involved in panel design and operation, from the electronics engineer to the airline flight instructor and government accident investigator, should be aware of the human factor principles applying to cockpit panels. Pilots, flight engineers, accident investigation officers, aircraft system designers, certifying authorities, maintenance engineers all, at some time, come face to face with the human factor aspects of flight station panel design.

Some are involved, protesting, because they cannot operate one switch without inadvertently actuating another, or because they cannot properly read an annunciator in the sunlight. Some are concerned with tight packaging of the electronics so that counters of sufficient size for the pilot to read can be installed. Others spend time retouching the paintwork of the panel because, in service, disturbing light leaks through the chipped edges, creating glare. Alle are engaged, wether they are aware of it or not, in activity associated with ergonomics of panel design.

Panel and display design technology is progressing. In time many of the problems discussed here, such as counter movement, will have a very special interest only. Nevertheless for many in the industry this day is still a long way off, and meanwhile it is necessary to optimise human factor aspects of current technology.

CURRENT DEFICIENCIES

But why should there still be deficiencies after all these years of experience? Why has the operator, the maintenance man, the flight instructor, the adapt to less than optimum panel design?

Panel design - like most cockpit design features - must be a compromise. A compromise because of space limitations, of availibility of suitable components, of electronic packaging requirements and of cost. Yet even allowing for these natural obstacles we could certainly do better, and the reason we have not done so in the past is largely owing to a lack of human factor orientation among those concerned with panel specification and design. And a lack of adequate, competent feed-back from the users to the manufacturers.

It may be necessary for an aircraft manufacturer to return a new panel many times to the vendor before it is finally found, with compromises, to be acceptable. This illustrates the problem we face in terms of both ergonomics and economics.

To some extent, the lack of human factor orientation can be explained by the fact that until recently there were no suitable training facilities anywhere in the world where those involved in the man-machine interface in transport aviation could get a short course in Human Factors directly applicable to their own industry and their own tasks. Although since 1972 different courses are available (1), and certain aircraft manufacturers have carried out some very limited internal programmes, it will be some time before all establishments have properly trained staff applying their knowledge in this vitally important aspect of design. And, of course, a prerequisite must be a recognition of the need for progress and a realisation that a better understandung of the applied technology of ergonomics is the key to such progress.

Whether designing a new panel, or evaluating one that has already been put together, a systematic and thorough human factor approach will pay off. It will ensure optimum operating efficiency and avoid costly changes to the hardware later in the design and development phase, or in service under pressure of operational necessity.

Here some general guidlines for this ergonomic exercise (06).

LOCATION AND ACCESSIBILITY: SIGHT AND REACH

Before beginning to consider details of the panel configuration, it is essential to review some aspects of the operational use to which the

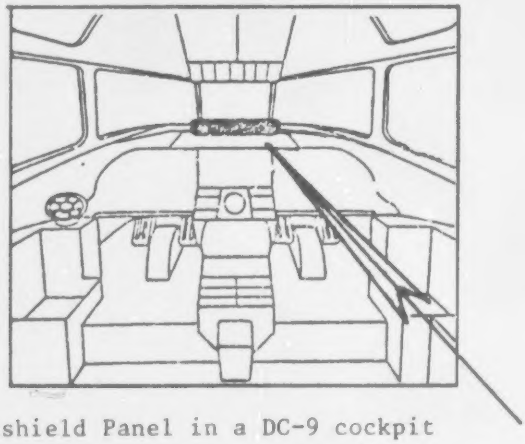


Fig. 01: Glareshield Panel in a DC-9 cockpit

panel controls or displays are to be applied. The location requirements for sight (displays) and reach (controls) may be quite different, though some controls - such as certain switches for which no other annunciation is provided - act as their own annunciator and must be seen as well reached. The glareshield panel (Fig. 01) is one example of a location that is rather suitable for both sight and reach, and, furthermore, accessible to both pilots.

It is imperative for the designer to know where the panel is to be located, so that he can take into account viewing angles and distances from the cockpit reference eye positions. This will influence the shape and size of panel cut-outs for instruments and readouts, mounting angles, location and size of nomenclature on the panel, and positioning of controls. Generally, 28-30 inches viewing distance is assumed for the main flight instruments panel ahead of each pilot.

LOCATION AND ACCESSIBILITY: PRIORITY FOR USE

In establishing the cockpit location of a panel, the aircraft manufacturer or airline evaluator should have taken into account operating procedures - who does what - and the priority of accessibility for each crew member. He should then have given specific instructions to the equipment vendor regarding the viewing angles required, and the direction from which switches or controls must be handled.

Some conflicts are bound to arise at this stage, because different airlines use different crew compositions; crew members may have different qualifications and a somewhat allocation of tasks. The common or basic panel location, and the panel design based upon it, is therefore not always optimum for all operators. This may be partcularly noticeable in connection with the ON/OFF position of switches, of which more later. However, there is a trend for the aircraft manufacturer to become more closely involved with operating procedures, and this works in favour of matching cockpit design to eventual airline crew operation.

Panels used in emergency procedures will need to be located in a position where they can always be reached, even if one crew member is temporarily absent from his station. Individual airlines may need other than a "basic" aircraft location of such panels to achieve this, taking into account their own crew task distribution policy.

LOCATION AND ACCESSIBILITY: INSTALLATION

Panels installed horizontally on the pedestal or consoles are vulnerable to damage from briefcases, tool boxes, spilled coffee and dirt. Long delicate control knobs and openings in or around the panel should therefore be avoided. The aircraft manufacturer and airline will ensure that cup holders (and ash trays) are kept well away from such panels, as the cost of spillage into panels can be a significant proportion of electronic panel maintenance. Special panel construction must be specified to minimise such damage risk.

Vertically mounted panels require special attention as switches may be difficult to operate if the pilot's hand is being bounced as a result of in-flight turbulence or taxying on a rough surface. It is adviceable to ensure the availability of some structure against which the pilot can steady his hand while making selections. Push-button switches may be used, altough this turbulence factor must again be kept in mind when arranging their size and proximity to each other.

Roof panels are normally too low and too close to the pilot's head for optimum operation. In particular, displays in this area are extremely difficult to design in such a way as the ensure proper visibility for both pilots, and their installation here should be avoided if possible. Counters recessed behind the panel face are very unsuitable on an overhead panel, as are small instrument dials, owing to parallax and bezel vision cut-off. For the more senior pilots today, "bifocal spectacles" may have become the name of the game-perhaps "trifocals", with the need to focus on the overhead panel, the main instrument panel, and the outside world. Incidentally, the question of head clearance has been studied by the SEA S-7 Committee and ARP 268E (01) is likely to be amended in the near future to establish new criteria.

At an early design stage the aircraft manufacturer should be aware of the proposed location of the panel in the cockpit, and should have noted any possible abstructions that may dictate the type of controls used and the location of the switches and displays.

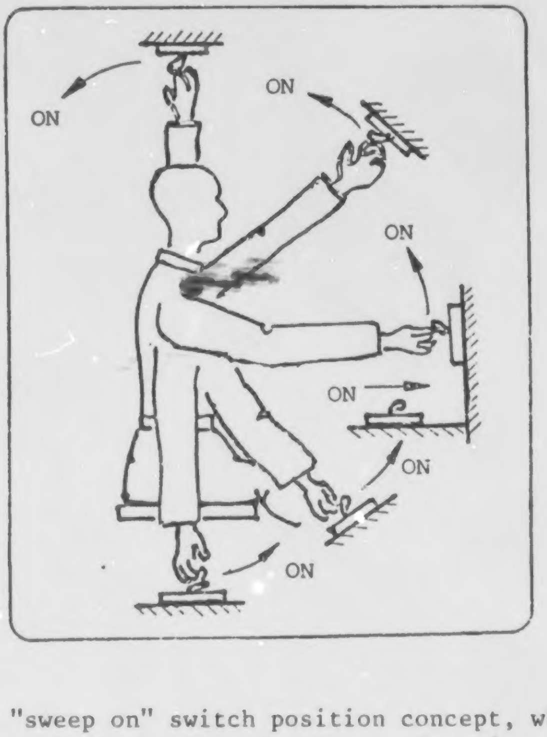


Fig. 02: The "sweep on" switch position concept, which avoids ambiguity on near vertical panels and allows relocation of panels from one cockpit location to another without disturbing conformity with the ON/OFF concept.

In addition, he should be aware of the incidence of reflections, and whether or not dirct sunlight can be expected to fall on the panel. Certainly he should specify an antireflective coating on all glass surfaces, and if direct sunlight can be expected to fall on the panel, this will rule out the use of many types of annunciator and dictate colour and brightness levels. For critical annunciators, such as those associated with Flight Guidance Modes, attention to this requirement is imperative, though too often neglected in current aircraft.

A fundamental choice that needs to be made is whether the nomenclature should be in upper case (capital) letters only, or in lower case as in this text. Here we must clearly differentiate between legibility - the ability to identify letters for what they are - and readability, which is the ability to recognize the group of letters as occurs during normal reading.

For speed of reading printed texts, lower case letters have been shown to be preferable owing to the better worldform. Many road signs are made this way. However, some experimental work has shown that for applications other than ordinary printing, upper case letters may be preferable as they seemed to result in fewer errors and faster recognition time.

Certainly, there will be no dispute that abbreviations such as ILS, HDG, VHF and so on should appear in capitals, though placards giving information as a text (e.g. caution placards) will no doubt remain a subject of debate.

The U.S. military specification (02) calls for upper case placarding only, and this policy appears to have become widely adopted. Yet, aircraft such as the british Hawker Siddeley Trident - whose cockpit was ergonomically quite advanced when conceived - used both upper and lower case; although now, owing to the need to standardise on one type upper case letters only have been adopted reluctantly on all new Hawker Siddeley aircraft. The International Organisation for Standardisation has published a document on this subject but in France, as elsewhere, this is only partly followed in the national standard.

The argument about the use of upper and lower case letters is not yet fully resolved, and we can expect it to be raised periodically in the future.

PANEL CHARACTERISTICS: LIGHTING

Panel lighting must be compatible with that of the rest of the cockpit. This is most likely to be assured if the panel is delivered by the aircraft manufacturer (Seller Furnished) because in new commercial aircraft the characteristics can then be expected to meet SEA ARP 1161 (03). When panels are Buyer or Customer Furnished, problems can be expected unless the airline has previously specifically required and checked that the panel lighting meets ARP 1161.

It is an astonishing and regrettable fact that in current aircraft, lighting specifications can differ not only between different manufacturers, but also between different aircraft models from the same manufacturer. The

83

availability of this ARP based on a wide measure of industry acceptance, may now reduce such discrepancies among future aircraft and provide not only better lighting qualita, but better equipment interchangeability between aircraft and more standard maintenance and manufacturing methods.

However, the introduction of newer techniques such as electroluminescence (EL), light emitting diodes (LED), and liquid crystals (LCD) may call for occasional up-dating of the publishing standards.

PANEL CHARACTERISTICS: NOMENCLATURE

There is, regrettably, no single international standard for the form of lettering on panals. For commercial aircraft, manufactureres ar free to select the style they prefer, bearing in mind that for practial reasons each customer cannot have his own particular choice. Letter and numeral height, width and spacing, and stroke width, can all vary according to the type selected.

PANEL CHARACTERISTICS: SWITCHES

When toggle switches are used, a basic decision on movement direction ON and OFF must be made. The two relevant philosophies are "forward on", which is self-explanatory, and "sweep on", with the "ON" position corresponding to sweeping the hand forward, up, and back overhead as shown in Fig. 02.

The "sweep on" concept has several advantages. It avoids the ambiguity that arises with panels close to the vertical (which way is forward?) and it permits free installation in and interchange between pedestal or console and overhead areas, without finishing up with a mixture of switches some of which are forward "ON" and some aft "ON". Standardisation of equipment is thus also facilitated, which holds down costs and enhances commonality - almost a holy word in aircraft specification today.

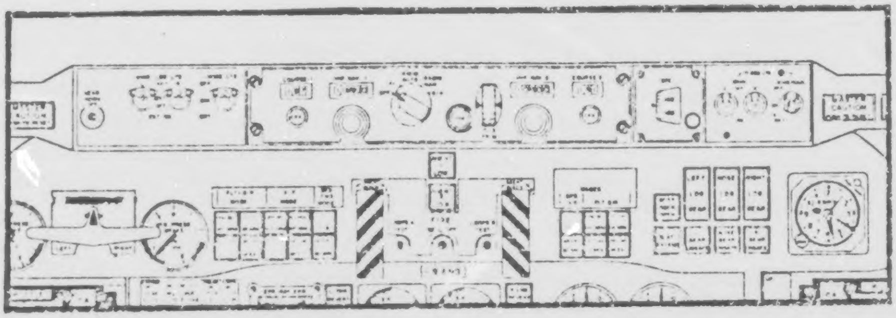


Fig. 03: DC-9 glareshield panel as a good example of the application of shape coding to control knobs

After having made a decision on the ON/OFF movement concept, the designer should ensure that inadvertent operation of one switch while operating another control is avoided, by careful location, leverlocking or guarding. He should check, too, that when viewed from the crew members reference eye positions the switch can cause no abscuration of panel nomenclature in any setting.

Coding of switches by shape or type is essential when several are installed close to each other (Fig. 03).

It should certainly be unnecessary for the crew member to have to read a panel placard to be shure of which switch he is operating.

The increasing use of push-button switches certainly eliminates the problem of obscuration, but it introduces others. The selected position of a push-button switch is far more difficult to recognize than of a toggle or rotary switch, and it will need to annunciate its own position, perhaps by lighting integral nomenclature as planned on Airbus 310. Brightness control of this annunciation is then required, taking into account both night and day conditions.

A further problem, referred to earlier, is that when operating push-buttons while taxying on rough surfaces or flying in turbulence there is nothing to "hang on to"; and, particularly when the panel is mounted vertically, there is a risk of actuating the wrong button. In such cases a support should be available with which to steady the hand while operating the switch.

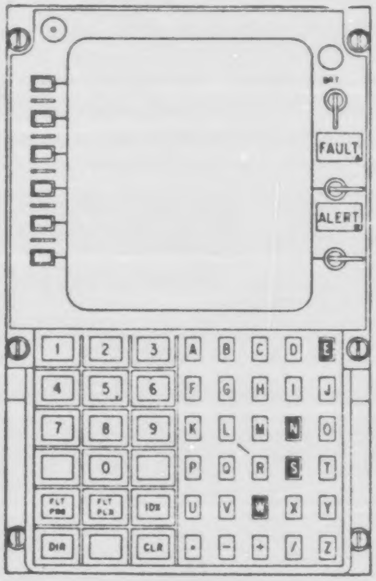


Fig. 04: The controls display unit of the Collins Area Navigational System. This illustrates ergonomic features of the keyboard such as the fence between the large buttons to minimize the risk of inadvertent operation, the layout of the alpha-numeric keyboard, colour coding to ease identification of N.E.S.W. on the alpha section.

PANEL CHARACTERISTICS: KEYBOARDS

Keyboards - collections of push-button switches - are now finding their way into modern aircraft, particulary for the control of navigation systems (Fig. 04), or computing performance data and AIDS; furthermore they will certainly be associated with Data Link. These present special problems.

Little attention had been paid to keyboard entry for aircraft used before 1967. Then however, stimulated no doubt by the forthcoming new families of widebodied aircraft and the progress made in solid state digital technology, Boeing, Douglas, Sperry and Collins among others became involved in studies and proposals. They compared the entry error rates, effects of turbulence, speed of entry and so on of keyboards compared with conventional knobs.

The layout of alpha-numeric keyboards has been widely researched. Perhaps the best-known early work was by Bell Telephone (04), but in aircraft, real estate is always scarce and optimisation of layout within severe space restraints is not easy. Fences between closely spaced pusbuttons may be necessary to reduce inadvertent actuations, and careful attention must be paid to determining the optimum force required to actuate the switch.

In all probability the cockpit operator is unlikely to have sufficient use of the panel to be able to operate the keyboard "blind", as does a touch-typist, even when all turning and entry requirements are made by centralised keyboards. It is desirable for the operator to be able to rest or steady his hand or wrist when operating a keyboard in an unstable environment such as an aircraft.

Colour coding can be used in white-lighted cockpits to enhance the layout. Even in red-lighted cockpits the alternate use of black on white and white on black can be helpful.

PANEL CHARACTERISTICS: KNOBS AND CONTROLS

A large variety of knob shapes and sizes is now available, some with more background of ergonomic research than others.

Knobs may be used for rotary switches, with fixed detents, or for continuos action controls. These all lend themselves to shape and colour coding, which should be applied to facilitate recognition. Different classes of knob were already categorised in 1954. The designer will need to properly review the published data before finally specifying the hardware. Obscuration of panel placarding (viewed from the reference eye position), shape, texture and colour, operating force, movement and ratio, graduation visibility during operation, parallax, design of graduation scale, detent characteristics - all need attention.

In general a knob should move clockwise to incrase the quantity. However, this general rule may be modified by what is known a Warwicks's Principle which, in this context, states that the indicator of a display is expected to move in the same direction as that point on the circumference of the rotary control which is closest to the display. The results of studies made in Australia (05) show the need to examine the control/display situation carefully in the light of the different - somitime conflicting - principles involved and that a wrong design can have a very large effect on performance.

Other examples of faulty design can be found, such as that seen on certain ATC Transponder panels (Fig. 05) where a "mirror image" instead of a "left-right" philosophy has been applied, resulting in frequent mis-selections even after years of use.

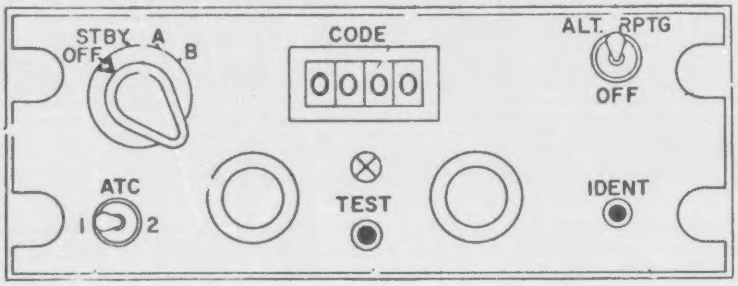


Fig. 05: Example of an ergonomically poorly designed panel. The left and right concentric knobs operate the left and right pair of digits respectively. However, the left large (lower) knob controls the L digit of the left pair while the right large (lower) knob controls the R digit of the right pair - a mirror image concept invariably causing selection errors.

Furthermore, digits 1, 2 and 4 move down when increasing, while digit 3 moves up. Even after many years of practice this panel still gives pilots control problems due to its poor ergonomic design.

Today there is little excluse for poor control knob design. Good surveys of the subject have been published and a vast amount of operational feedback has been made to equipment vendors. Yet errors continue to be made; both designer and evaluator need to do their homework properly in this area.

PANEL CHARACTERISTICS: COUNTERS, DRUMS AND DIGITAL READ-OUTS

Alpha numeric read-outs may be settable by the crew member (e.g. speed command) or be simply a display of system or environmental conditioning (e.g. SAT).

They may be mechanical, elekctro-mechanical, or electronic. Many of the human factor principles apply to some extent to all such displays, so these brief notes will remain rather generalised.

Standardisation in counter, drum and digital read-outs is only just beginning to appear. Currently not only do counters of different displays move in different directions to increase the value, but sometimes even separate digits in a single read-out (Fig. 05) move in opposite directions. For magnetic wheels, which have a rapid transfer action, this is not very significant. However, for other cases,particulary where the read-out is set manually by the pilot, this can be an annoying feature which reduces efficiency in the execution of cockpit tasks.

When considering the correct direction of movement there may be different principles applying between a drum or counter on, for example, an altimeter which has some spatial relationship, and other read-outs, say, of SAT or TAS. The altimeter drum or counter should move down as the altitude increases, and there now seems to be a trend to design all digital read-outs this way. Publications give conflicting advice on this question.

In marking the choice between a rapid-transfer-action counter (e.g. magnetic wheel) and a continously rotating drum, consideration has to be given to the possivle need for rate information, which is reasonable well available (at slower rotation speeds) with the drum, but rather poor with the rapid-action counter.

It is certainly undesirable for counter digits to be less than 3/16 inch high, and where possible a minimum of 1/4 inch should be the target. Read-outs at one cockpit station which have to be read from another station may need to be larger than this.

A fast slewing facility is often needed for proper manual setting of digital read-outs, to take advantage of the accuracy and range desired and to avoid problems in setting the gain. This feature is frequently poorly designed ergonomically.

We have referred to mechanical and electro-mechanical read-outs, but we are now seeing a trend to electronic read-outs, using techniques such as CRT, EL, LED, LCD, and various projected and segmented annunciations. While the scope of this report does not permit a detailed analysis of the human factors relating to these techniques, brightness, contrast, colour, viewing angle, and alpha-numeric format all need ergonomic attention. Replacement of the mechanical and electro-mechanical read-outs by these new electronic displays will make arguments on the question of counter movement of academic interest only.

CONCLUSION

From this survey it will be apparent that without proper attention to Human Factors there are many deficiencies that can remain in the operational hardware.

It is becoming increasingly recognised that failure to properly match the hardware to the characteristics of man induces stress into this complex man-machine system, leading to efficiency and contributing to potential breakdown.

A systematic and educated application of sound ergonomic principles in the design stage, during development and in operational service is therefore essential. And the earlier in the design process these principles can be applied, the greater is the probability of a successful outcome and the lower is the cost of achieving this

REFERENCES

(01) SAE ARP 268E, Location and Actuation of Flight Deck Controls for Commercial Type Transport Aircraft.

(02) MIL-M-18012B, Marking Aircrew Station Displays, Design and Configuration of

(03) SEA ARP 1161, Crew Station Lighting, Commercial Aircraft

(04) L. DEININGER Human Factors Engineering Studies of the Design and Use of Push Button Telephone Sets, Bell System Technical Journal, Volume 39, 1960

(05) BREBNER J. and SANDOW B.: Direction-of-turn stereotypes-conflict an control, Applied Ergonomics, March 1976

(06) F. HAWKINS Some Ergonomic Aspects of Cockpit Panel Design for Airline Aircraft, SAN 437

(07) APPLIED ERGONOMICS, IPS Science Technology Press LTd, 32 High Street, Guildford, Surrey, England

(08) A. DOVALIL Ergonomische Aspekte im System Pilot-Flugzeug-Umwelt. diploma lecture note Technical University Vienna, 1979

THE AUTORS

Hellfried AUBAUER, born 1941, graduated and made his PhD in Mechanical Engineering at the Technical University of Vienna. Interested in flying, he joined AUSTRIAN AIRLINES, where he is active as Captain, Flight Instructor, Technical Pilot and Accident Investigation Officer. He is lecturning Flight Operations Technology at the University of Vienna. Since 1979 he is an active member of the AUSTRIAN FLIGHT SAFETY BOARD.

Walter SPERR, born 1937, BS in Mechanical Engineering and BS in Business Administration, made his PhD in Technical Sciences at the Technical University of Vienna. Active in Structures Engineering, he held research contracts with Aircraft Industries in France, West Germany and Netherlands. Since 1972 he is Independent Consultant in Structural Design and Failure Analysis of Lightweight Constructions, lecturer at the University of Vienna, and member of the AUSTRIAN FLIGHT SAFETY BOARD. He has published more than 40 papers and one book in Failure Analysis.

THE ORGANIZATION

The AUSTRIAN FLIGHT SAFETY BOARD, founded in 1979, is a scientific and technical organization formed for the "research, development and promotion of procedures and possibilities to prevent aircraft accidents".

A COMPARISON OF TRACKING WITH VISUAL AND KINESTHETIC-TACTUAL DISPLAYS

R. J. JAGACINSKI, J. M. FLACH, AND R. D. GILSON

Human Performance Center and Aviation Psychology Laboratory
The Ohio State University
Columbus, Ohio

ABSTRACT

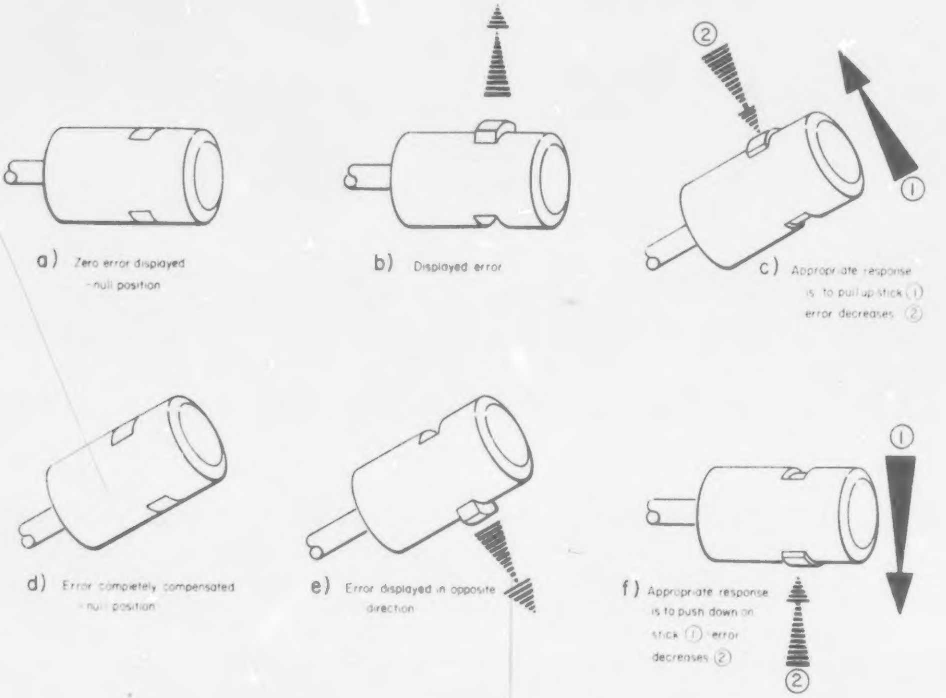
Recent research on manual tracking with a kinesthetic-tactual (KT) display suggests that under appropriate conditions it may be an effective means of providing visual workload relief. In order to better understand how KT tracking differs from visual tracking, both a critical tracking task and stationary single-axis tracking tasks were conducted with and without velocity quickening. On the critical tracking task, the visual displays were superior; however, the KT quickened display was approximately equal to the visual unquickened display. Mean squared error scores in the stationary tracking tasks for the visual and KT displays were approximately equal in the quickened conditions, and the describing functions were very similar. In the unquickened conditions, the visual display was superior. Subjects using the unquickened KT display exhibited a low frequency lead-lag that may be related to sensory adaptation.

INTRODUCTION

The question is often asked, Are the aircraft of the future to use only automated control? Are we at the point of using only pushbutton inputs into intelligent flight control circuitry--without manual control at all--the ultimate fly by wire? Not today at least--even the space shuttle has a traditional control yoke (although the throttles are non-existent in this brick-like glider). Thus manual tracking control still has its place in today's current and envisioned aircraft.

How do we provide information in the cockpit? Presently via two traditional senses--the eyes and ears--overburdened as they may be with the barrage of information in this high speed, complex and heavily used mode of travel. The other modalities of information input are, of course, the senses of smell, taste, and touch. While the first two (smell and taste) are currently impractical for communication or control, some little research has been devoted to touch as an informational channel. The bulk of tactile research has centered on language for the blind (e.g. Braille); however, some research during the last decade has addressed the use of touch (e.g., electrocutaneous, vibratory) as a means for compensatory tracking in aircraft (e.g., attitude, glideslope). One of the more successful techniques has been a kinesthetic-tactual (KT) display (Figure 1) invented by Professor Robert Fenton of The Ohio State University

Department of Electrical Engineering (Fenton, 1966; Fenton and Montano, 1968). Simulator and inflight testing (Gilson and Fenton, 1974; Gilson, Dunn, and Sun, 1977) suggested considerable promise for this KT display as a substitute for visual displays of flight control information, particularly during periods of visual distraction. By presenting information to multiple modalities, it may be possible to increase a pilot's overall workload capability. For example, a recent laboratory investigation (Burke, Gilson, and Jagacinski, 1980) of a two-handed tracking task has shown an advantage for a combination of KT and visual displays over a two-dimensional visual display.


a) Zero error displayed
- null position

b) Displayed error

c) Appropriate response
is to pull up stick ①
error decreases ②

d) Error completely compensated
- null position

e) Error displayed in opposite
direction

f) Appropriate response
is to push down on
stick ① - error
decreases ②

Figure 1. Display-response relationship for the KT display. (Copyright 1979, Human Factors, Vol. 21, p. 80)

The current study is an attempt to analyze the characteristics of the KT display and the operator in order to measure describing functions that may generalize across displays. Subjects were first practiced on a critical tracking task (Jex, McDonnell, and Phatak, 1966) using KT and visual displays with and without velocity quickening. Subjects were then transferred to a stationary tracking task involving either a rate controller (single integrator plant) or a first-order unstable system, and linear describing functions were derived from their performance. The direct and practical benefit of this engineering/psychological approach is to point out favorable control tasks for implementing the tactual display, point out similarities and differences in the operator's use of visual versus tactual displays, and perhaps give some insights as to display characteristic optimization for future designs.

METHOD

Subjects

Thirty-two right-handed undergraduates were chosen on the basis of a pretest described below and then were assigned to one of eight experimental conditions. Subjects either received credit for an undergraduate laboratory requirement for up to four experimental sessions or were paid $2.50 per session.

Apparatus

Visual and KT displays were used with and without velocity quickening. The visual display consisted of a 1 cm long horizontal line that moved vertically on a Tektronix Type 602 CRT display. The center of the display was indicated by a 1 mm x 18 mm strip of yellow tape. Full scale was ±4 cm, which corresponded to ±3.76° of visual angle. The KT display was built into the cylindrical handle of the control stick, and consisted of a servo-controlled solid rectangular section (1.25 cm x 2.2 cm x 3.8 cm long) sliding in and out of the handle (Figure 1). Excursions from the flush surface of the control handle indicated the direction and magnitude of system error. Full scale was ±1 cm. For the quickened displays the ratio of position to velocity was 1:1.

The control stick was an isotonic lever arm that was 52 cm long from the pivot point to the center of the KT display. The control stick moved in a vertical plane at the left side of the seated subject and resembled a helicopter collective control. The range of angular travel was ±10°, with 20° above horizontal representing the neutral control position. The control stick was counterbalanced and had a nominal level of friction so that no force was necessary to maintain any angular position. Approximately 250 gm of force was necessary to move the stick from an initially stationary position.

Pretest

Eight groups of eight to ten subjects each were pretested on a critical tracking task (Jex, et al., 1966) using an unquickened visual display and a small isotonic control stick that required wrist movements performed with the right hand. The four subjects in each group with the best performance over the last 15 out of 45 trials were chosen to continue in the experiment.

Critical Tracking

The eight groups of four subjects performed 45 critical tracking trials per day for seven days. Subjects sat in a 2.1 x 1.5 m room with low ambient light for the visual displays and in total darkness for the KT displays. White noise heard through a headset masked auditory cues from the KT display servo-drive. On each trial subjects used their left hands to manipulate a control stick which resembled a helicopter collective. The control stick gain was .56° of visual angle for the visual display and .15 cm for the KT display per 1° of control stick movement. The subjects' task was to maintain control of a first order unstable system as its time constant was progressively shortened over the course of a trial. No input signal was provided because the subjects' manual unsteadiness was sufficient to perturb the unstable system. The error signal, i.e. any deviation from zero output, was presented on one of four displays: two groups used an unquickened KT display; two groups used a quickened KT display; two groups used an unquickened visual display; and two groups used a quickened visual display. For the quickened conditions, the display gain was halved in order to keep the display range comparable to the unquickened conditions. For the KT display conditions a red warning light visible to both the subjects and experimenter was turned on whenever subjects held the KT display tightly enough to impede its movement.

The inverse of the time constant of the first order system, λ, was linearly increased over the course of a trial at .05 r/s^2 until the subjects allowed the error signal to reach full scale. At that instant the trial was terminated, and the value of the inverse time constant, referred to as the critical root, λ_c (Jex, et al., 1966), was displayed to the subjects and recorded as the performance measure. For the quickened displays it was the true system error, rather than the displayed quickened error, which was used to determine the end of a trial. The value of λ at the start of a trial was chosen to be approximately 1.5 r/s less than the value of λ_c on the immediately preceding trials, so that each trial lasted approximately 30 s. The trials were administered in three blocks of 15, with a two-minute break between blocks.

Stationary Tracking

Following the seven days of critical tracking, subjects were transferred to three days of stationary tracking. Subjects used compensatory displays of the same types used in their critical tracking tasks, with the exception that the red warning light was omitted from the KT display conditions, and the display gain for the quickened conditions was not halved. The system dynamics were either a single integrator, 1.5/s, or a first-order unstable system, 3.0/(s-1). The static gain of the control stick was 1.41° of visual angle or .375 cm of KT display displacement per 1° of control stick movement for the 1.5/s system. The static gain was twice these values for the 3.0/(s-1) system. The input signal consisted of a sum of nine sinewaves with the amplitudes of the three lowest frequency sinewaves (.35, .73, 1.08 r/s) five times greater than the amplitudes of the other sinewaves. Subjects performed eight three-minute trials per day for three days. After each trial subjects were told their integrated squared error and received a one-minute break before the next trial. A $5 bonus was given to the subject in each display group having the lowest error score on the last day of the experiment.

RESULTS

Critical Tracking

The median value of λ_c was calculated for each block of trials for each subject on the seventh day of critical tracking. The mean of these medians for each group of subjects is shown in Figure 2. An analysis of variance performed on the mean λ_c scores revealed statistically significant main effects of sensory modality and quickening ($p < .01$) with no significant interactions. There was also no effect of the system dynamics to which the subjects were assigned for the subsequent stationary tracking. These results closely resemble those obtained by Jagacinski, Miller, and Gilson (1979) with the exception that performance with the KT displays is noticeably better given this extensive amount of practice. Based on these results, one would expect the visual displays to be better than or equal to the KT displays for the stationary tracking.

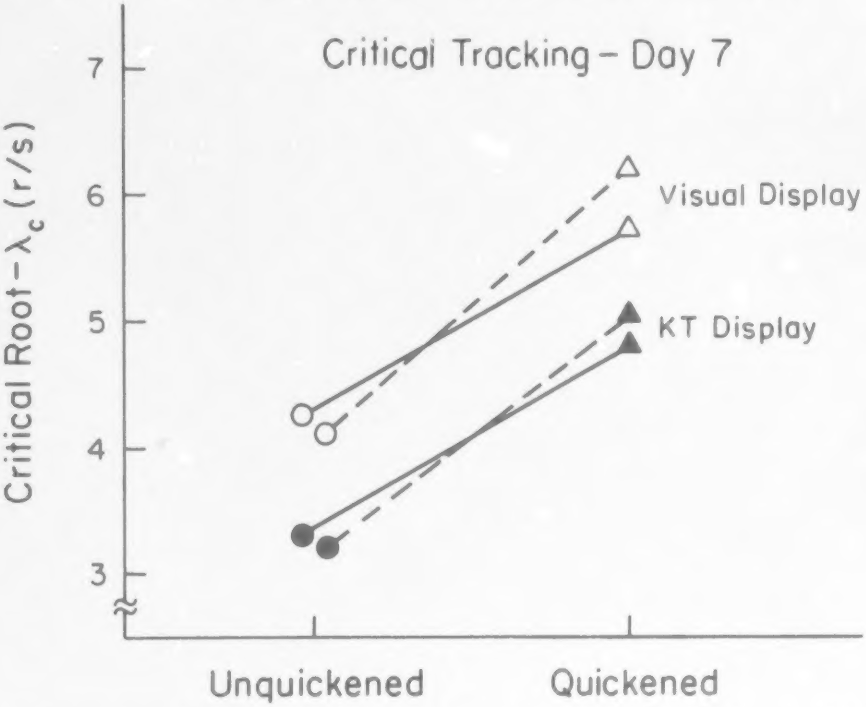


Figure 2. Critical tracking scores for eight groups of four subjects. Groups connected by dashed and solid lines respectively transferred to stationary tracking with system dynamics 1.5/s and 3.0/(s-1).

Stationary Tracking

The mean of each subject's mean squared error scores normalized by mean squared input for the third day of stationary tracking are shown in Figure 3. One subject who used an unquickened KT display was unable to perform the stationary tracking task with 3.0/(s-1) dynamics, and this subject is omitted from Figure 3 and all subsequent analyses.

Quickened Displays. For the quickened display conditions, individual differences among subjects were relatively small, and differences between the visual and KT displays were small. An analysis of variance revealed no statistically significant main effects or interactions ($p > .10$). A describing function was derived for each subject by calculating Φ_{ic}/Φ_{ie} at each of the nine input frequencies (McRuer, Graham, Krendel, and Reisener, 1965). Φ_{ie} is the cross spectral density of input and displayed error, and Φ_{ic} is the cross spectral density of input and control stick position. The mean amplitude ratio and mean phase shift were calculated from four trials of Day 3 of stationary tracking. These means are displayed as circles in Figure 4 for the four subjects having the lowest mean squared error in their respective display conditions.

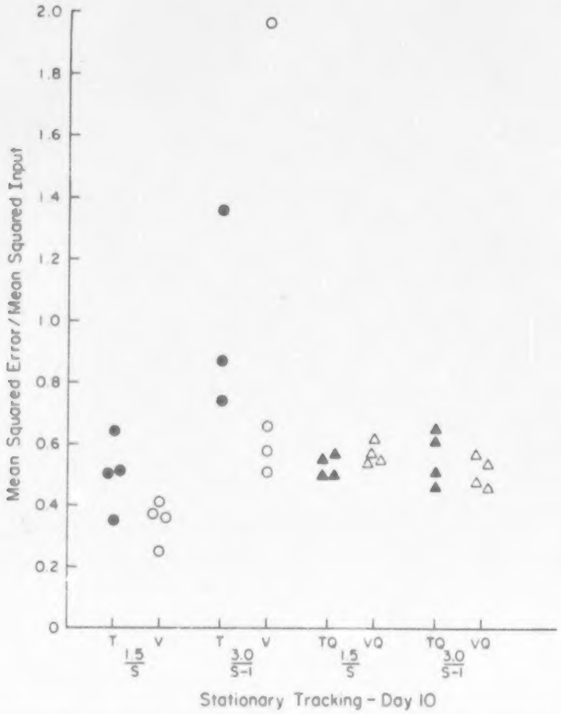


Figure 3. Mean squared error normalized by mean squared input for thirty-one individual subjects. The symbols represent the same display conditions as in Figure 2.

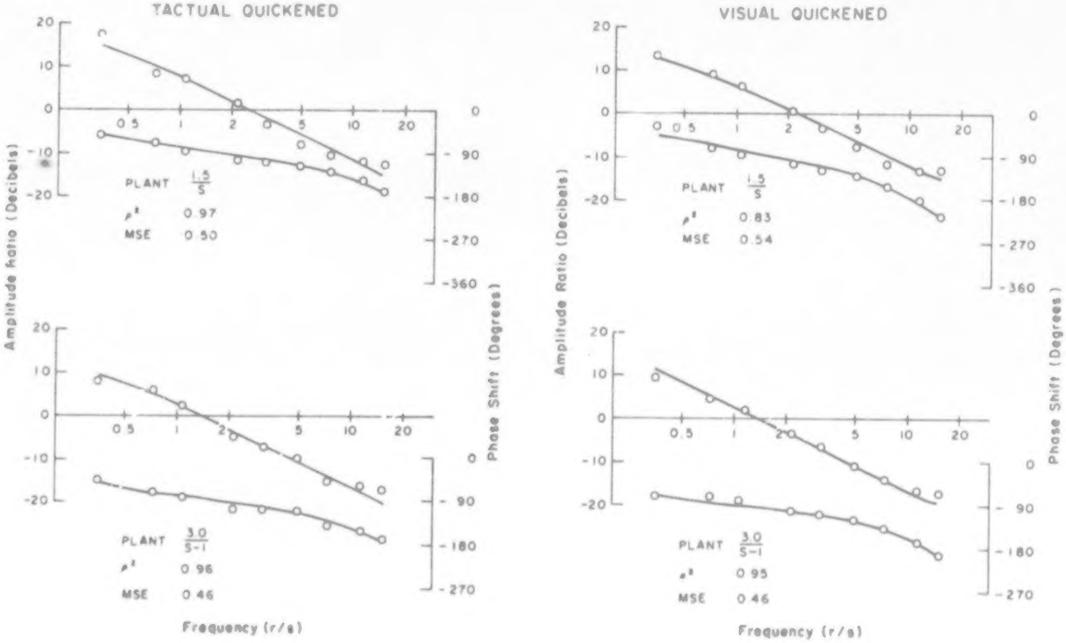


Figure 4. Linear transfer functions for the subjects with the lowest mean squared error in each of four quickened display conditions. The circles indicate the data points, and the solid lines represent analytic approximations consisting of a low frequency lag, a high frequency lead, a gain, and a time delay.

These describing functions are well approximated by a low frequency lag, a high frequency lead, a gain, and a time delay. Values of these parameters were chosen to maximize the proportion of variance accounted for among the amplitude ratios plus the proportion of variance accounted for among the phase shifts. For each subject in Figure 4, the linear transfer function resulting from this parameter search is represented as a solid line. For four of the sixteen subjects the high frequency lead was effectively absent. These four subjects were distributed across three experimental conditions, and all of these subjects had the highest or next to highest mean squared error in their respective groups. Across the quickened conditions, the visual and KT describing functions were very similar.

The proportion of variance of each subject's control that was linearly correlated with the input was estimated as $\rho^2 = \sum_{n=1}^{9} \frac{1}{2} I^2(\omega_n)[\Phi_{ic}(\omega_n)/\Phi_{ii}(\omega_n)]^2 / \overline{C^2}$ (McRuer, et al., 1965). $I(\omega_n)$ is the amplitude of the input sinewave at frequency ω_n, $\Phi_{ic}(\omega_n)$ is the cross spectral density of input and control at this input frequency, $\Phi_{ii}(\omega_n)$ is the auto power spectral density of the input at this frequency, and $\overline{C^2}$ is the mean squared control. The mean value of ρ^2 averaged across subjects ranged from .86 to .94 for the four quickened display conditions.

The describing function for the KT display was calculated from the command signal to the display and the display response as indicated by a follower-potentiometer coupled to the slide. This describing function was well approximated by a first-order lag with a break frequency of 18.5 r/s and a time delay of .01 s. The display lag is not included in the describing functions for the KT subjects in Figure 4.

Unquickened Displays. For the unquickened display conditions, there were large individual differences in mean squared error with the 3.0/(s-1) dynamics, which precluded using an analysis of variance. Therefore, a t-test was used for the 1.5/s system, and a non-parametric test was used for the highly variable data with the 3.0/(s-1) system. A t-test comparing the visual and KT displays with the 1.5/s dynamics indicated a tendency for the visual display to yield lower mean squared error ($t = 2.2$, $p < .05$, one-tailed). For the 3.0/(s-1) dynamics, if one ignores the single outlier in the visual condition, then all three of the remaining visual subjects performed better than the three KT subjects. A Mann-Whitney U-test indicates that this result is also statistically significant ($U = 0$, $p = .05$, one tailed).

The median amount of control used by the groups of subjects with the unquickened displays ranged from 3.32 to 6.76 root mean squared degrees of stick movement. This amount of control is considerably greater than for the groups of subjects with the quickened displays, whose median root mean squared control ranged from 1.49 to 2.50 degrees. For the unquickened displays, the proportion of variance in the control movements linearly correlated with the input frequencies was considerably less than for the quickened displays. For the 1.5/s dynamics, the mean value of ρ^2 was .69 for the unquickened visual display and .42 for the unquickened KT display. Inspection of power spectra of the control movements of subjects using the KT display revealed from one to three strong peaks occurring at non-input frequencies in a region from approximately 3 to 7 r/s. Corresponding peaks were sometimes present, but considerably smaller in the control spectra of subjects using the visual display. This control activity may represent relay-like control superimposed on more nearly linear tracking behavior. For the 3.0/(s-1) dynamics, the mean value of ρ^2 was .68 for the unquickened visual display and .69 for the unquickened KT display. There was some evidence of peaks occurring in the control spectra at non-input frequencies for this KT group, but these peaks were much smaller than with the 1.5/s dynamics.

Describing functions for the subjects with the lowest mean squared error in their respective unquickened display conditions are shown as circles in Figure 5. The data are well approximated by linear transfer functions consisting of a low frequency lag and lead, a high frequency second-order lag, a gain, and a time delay. Subjects using the KT display generally exhibited less phase lag at the three lowest frequencies. This difference from the visual display conditions can be modeled as an additional lead-lag for the KT group. The describing function for the electromechanical KT display itself was the same as in the quickened conditions. The KT display lag is not included in the describing functions of the KT subjects in Figure 5.

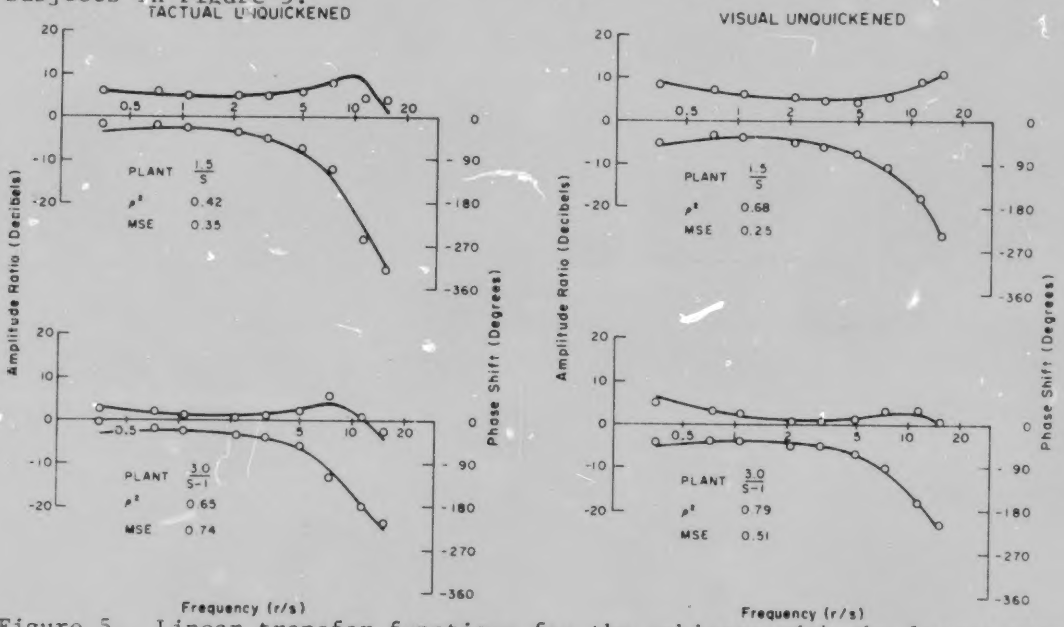


Figure 5. Linear transfer functions for the subjects with the lowest mean squared error in each of four unquickened display conditions. The circles indicate the data points, and the solid lines represent analytic approximations consisting of a low frequency lag and lead, a high frequency second-order lag, a gain, and a time delay.

DISCUSSION

The quickened visual displays yielded better performance than the quickened KT displays in the critical task, but the two displays yielded approximately equal performance in the stationary tracking tasks. The results of Burke et al. (1980) suggest that the difference in critical task performance with the quickened displays is primarily due to the electro-mechanical lag of the KT servo-drive. It may be that with increasingly unstable systems or with input signals of higher bandwidth the stationary tracking results would also reflect the ordering found in the critical task. However, tracking simulations based on the describing functions of subjects with the KT lag removed from the system suggest that the electro-mechanical lag contributed little to the obtained error scores in the quickened KT conditions. These stationary tracking tasks involved relatively little control movement, the display moved relatively rapidly due to the quickening, and subjects' behavior was strongly linearly correlated with the input signal. The describing functions were well approximated by a low frequency lag, a high frequency lead, and a time delay, and there was surprisingly little difference in this pattern between modalities.

For the unquickened displays, subjects performed better with the visual display in the critical task and in the stationary tracking tasks as well. Tracking simulations suggest that the electro-mechanical KT display lag contributed little to the error scores for the 1.5/s system, but may account for approximately half the difference in error scores between KT and visual unquickened displays with the 3.0/(s-1) system. This stationary tracking involved a good deal of control movement, the displays moved more slowly than the quickened displays, and subjects' behavior was much less linearly correlated with the input. The describing functions were well approximated by a low frequency lag and lead, a high frequency second-order lag, and a time delay. Differences between the visual and KT describing functions at low frequencies were well approximated by an additional lead-lag for the KT describing functions. This pattern may be indicative of a kind of rapid sensory adaptation (Milsum, 1966) in the kinesthetic-tactual sensory system that is not present in the visual system. In other words, the KT sensory system may be relatively more sensitive to velocity than position stimuli, and this factor probably contributed to the differences in error scores with the unquickened displays.

A display technique that might enhance velocity cues and/or reduce non-linearities in the KT sensory system is to add a high frequency, low amplitude vibration to the KT display. The frequency would have to be high enough that this signal would not be confused with error correlated with the input signal. Whether this technique does in fact improve performance with the unquickened KT displays remains to be tested.

In summary, the present experiments have shown that with quickened displays and a low input bandwidth tracking performance is approximately equivalent with visual and KT displays for a single integrator and first-order unstable systems. With unquickened displays, the visual modality is superior. This superiority may be due to the KT modality being relatively more sensitive to velocity rather than position cues as well as to the electro-mechanical lag in the KT display.

ACKNOWLEDGMENTS

This research was sponsored in part by the U.S. Army Air Mobility Research and Development Laboratory, Moffett Field, California, and in part by the Avionics Laboratory, Electronics Command, Ft. Monmouth, New Jersey, and was monitored by Richard S. Dunn through NASA-Ames Grant NSG-2179. The authors wish to thank Dwight Miller, Christopher Williams, Mary Malone, and Shirlee Conley for their assistance in the data analysis.

REFERENCES

Burke, M. W., Gilson, R. D., and Jagacinski, R. J. Multi-modal information processing for visual workload relief. Ergonomics, 1980, 23, 961-975.

Fenton, R. E. An improved man-machine interface for the driver-vehicle system. IEEE Transactions on Human Factors in Electronics, 1966, 7, 150-157.

Fenton, R. E., and Montano, W. B. An intervehicular spacing display for improved car-following performance. IEEE Transactions on Man-Machine Systems, 1968, 9, 29-35.

Gilson, R. D., Dunn, R. S., and Sun, P. A kinesthetic-tactual display concept for helicopter-pilot workload reduction. Paper to the 33rd Annual National Forum of the American Helicopter Society, Washington, D.C., 1977.

Gilson, R. D., and Fenton, R. E. Kinesthetic-tactual information presentations--Inflight studies. IEEE Transaction on Systems, Man, and Cybernetics, 1974, 4, 81-90.

Jagacinski, R. J., Miller, D. P., and Gilson, R. D. A comparison of kinesthetic-tactual and visual displays via a critical tracking task. Human Factors, 1979, 21, 79-86.

Jex, H. R., McDonnell, J. D., and Phatak, A. V. A critical tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 138-145.

McRuer, D., Graham, D., Krendel, E., and Reisener, W., Jr. Human pilot dynamics in compensatory systems. Air Force Flight Dynamics Laboratory Technical Report Nr. AFFDL-TR-65-15, Wright-Patterson Air Force Base, Ohio, July, 1965.

Milsum, J. H. Biological control systems analysis. New York: McGraw-Hill, 1966.

GENERAL AVIATION COCKPIT DESIGN FEATURES RELATED TO INADVERTENT LANDING GEAR RETRACTION ACCIDENTS

Alan Diehl
Former Senior Air Safety Investigator
National Transportation Safety Board
Washington, D.C.

SYNOPSIS

A detailed review was made of all National Transportation Safety Board (NTSB) files of inadvertent landing gear retraction accidents occurring to general aviation aircraft in the U.S. from 1975 to 1978. The data indicated that two particular types of airplanes were involved in the majority of these accidents although they comprised only one-quarter of the active light aircraft with retractable landing gears. Pilot comments and human engineering evaluations of contemporary light aircraft cockpits revealed that these two particular aircraft types have four design features which should tend to increase the probability of inadvertent landing gear retraction accidents.

INTRODUCTION

This paper is based upon a recent NTSB special investigation (SR-80-1) made by the author while employed by that organization. 1/ An indepth examination was performed on the files for every inadvertent landing-gear retraction accident occurring in the U.S. between 1975 and 1978. This effort revealed that while such accidents invariably involved aircraft damage, they resulted in little or no injury to the occupants. But, unlike the situation in many other kinds of "pilot error" accidents, these untramautized pilots for the most part were willing and able to state the reasons behind their mistakes. The files indicated that these accidents typically occurred because the pilot was attempting to put the flaps control "up" after landing, but moved the landing gear control upwards instead. The inadvertent movement of the landing gear control was often attributed to the pilot's being more accustomed to flying aircraft in which these two controls were in exactly opposite locations.

These data also indicated that two aircraft types, the Beech "Bonanza" (Models 33, 35, and 36), and the Beech "Baron" (Models 55, 56, 58, and 95) were involved in 61 percent of the inadvertent landing gear retraction accidents which occurred during that four year period. The Bonanza and Baron, 2/ however, constituted only about 25 percent of the active light aircraft fleet with retractable landing gears. These results were similar to those reported in an earlier study (NTSB,1969). That report concluded that the early Bonanzas, while comprising only 22 percent of the fleet with retractable landing gear, accounted for 48 percent of the inadvertent gear retraction accidents.

Comparing the details of Bonanza and Baron's cockpit features to those of other contemporary light aircraft revealed they differed in certain ways from most other contemporary light aircraft, e.g. landing gear and flap control locations.

1/ The full report is available from the NTSB, Washington, D.C. 20594.

2/ These two aircraft were also marketed under the names "Debonair" and "Travel Air," respectively.

Such human engineering problems may result largely from the fact that their basic cockpit control arrangement is approximately 35 years old. Since these aircraft were originally certificated, the influence of good design in preventing human error has been more clearly shown. This report examines how certain cockpit design features appear to have generated the relatively high rates 3/ of inadvertent gear retraction accidents in these two airplanes. In addition, it will show how these deficiencies apparently have contributed to accidents in other types of aircraft because the pilots were more familiar with the nonstandard arrangement of the Bonanzas and Barons.

STATISTICS

The FAA records for 1978 indicate that the various Beechcraft Bonanza models comprised 9,430 aircraft in a fleet of approximately 31,500 active single-engine aircraft with retractable landing gear. NTSB data indicate that from 1975 through 1978, these Bonanzas were involved in 16 of the 24 inadvertent gear retraction accidents. (See Table 1.) Thus the Bonanzas comprised only about 30 percent of the single-engine aircraft fleet with retractable gears, but they were involved in 67 percent of such accidents.

The 1978 FAA records showed that the various Beechcraft Baron models comprised 3,441 of the approximately 21,000 active reciprocating engine light twins. During the 1975 to 1978 period, NTSB records indicated that the Barons suffered 21 of the 39 inadvertent gear retraction accidents. (See Table 2.) Thus the Barons comprised only about 16 percent of the light-twin fleet, but they were involved in 54 percent of the accidents of this type.

The data indicate that the Bonanzas and Barons accident rates are generally several times as great as those of most other similar contemporary light aircraft. Figure 1 graphically illustrates these facts. For instance, the differences in the rates of occurrence of inadvertent landing gear retraction accidents can be seen by comparing the Bonanza with a similar aircraft, the Cessna 210. The 4,741 Cessna 210's, which comprised 15 percent of the single-engine retractable gear fleet in 1978, were only involved in 4 percent (1 accident) of such mishaps occurring to single-engine aircraft from 1975 to 1978. In contrast, the Bonanzas, comprising about 30 percent of the fleet, experienced 67 percent of these accidents (21 accidents) -- an accident rate about 10 times as high as that of the Cessna 210. A standard statistical test for comparing differences in proportions produced a "z" equal to 2.5 which indicated that this accident rate was significantly worse for the Bonanza at the .01 level of confidence (Spiegel, 1961).

Likewise, the accident rate of the Baron can be compared to a similar light twin, the Piper P-23 Aztec, 4/. The 3,459 active PA-23's comprised about

---

3/ These rates were derived for each type aircraft by dividing the number of inadvertent landing gear retraction accidents by the estimated number of those aircraft which were active.

4/ The early models of the PA-23 were marketed under the name "Apache."

101

TABLE 1

RETRACTABLE LANDING GEAR ACCIDENTS: SINGLE ENGINE AIRCRAFT

Model	Date	Location	Total Pilot Hours in Accident Involved Model	Total Pilot Hours in All Makes and Models	Pilot Admitted Confusing Flaps with Landing Gear	Pilot Stated a Familiarity for a Reversed Arrangement of Gear and Flaps
BEECHCRAFT BONANZA						
BE-33	1/19/75	Elko, NV	120	3564	x	x
BE-35	6/15/75	West Mifflin, PA	50	630	x	—
BE-35	6/19/75	Jackson, OH	220	1056	x	—
BE-35	9/2/75	Minneapolis, MN	111	293	x	—
BE-35	9/7/75	Laramie, WY	870	1173	x	—
BE-36	3/19/76	Camdenton, MO	525	1656	—	—
BE-35	3/21/76	St. Joseph, MO	131	2214	—	—
BE-35	4/13/76	Jacksonville, FL	16	2950	x	—
BE-35	7/3/76	San Carlos, CA	183	406	—	—
BE-35	5/4/77	Jacksonville, FL	30	416	x	—
BE-35	6/4/77	Bessemer, AL	[illegible]	4426	x	—
BE-35	8/5/77	Dallas, TX	51	136	—	—
BE-33	8/28/77	Keystone Heights, FL	7	1330	x	—
BE-36	8/30/77	Keystone Heights, FL	228	5176	—	—
BE-35	10/16/77	San Jose, CA	3	337	—	—
BE-35	7/6/76	Meadville, PA	127	368	x	—
PIPER						
PA-24	5/10/76	Greenwood, SC	1220	1240	—	—
PA-24	6/12/76	Muleshoe, TX	665	665	—	—
PA-24	7/13/76	Columbia, SC	550	800	—	—
PA-24	11/23/76	Birmingham, AL	20	4739	—	—
MOONEY						
M-20	7/[illegible]/76	Las Vegas, NV	11	887	—	—
M-20	5/20/78	Westerly, RI	187	405	—	—
M-20	7/21/78	Atlanta, GA	430	1075	—	—
CESSNA						
C-210	5/19/77	Amarillo, TX	40	3000	x	—

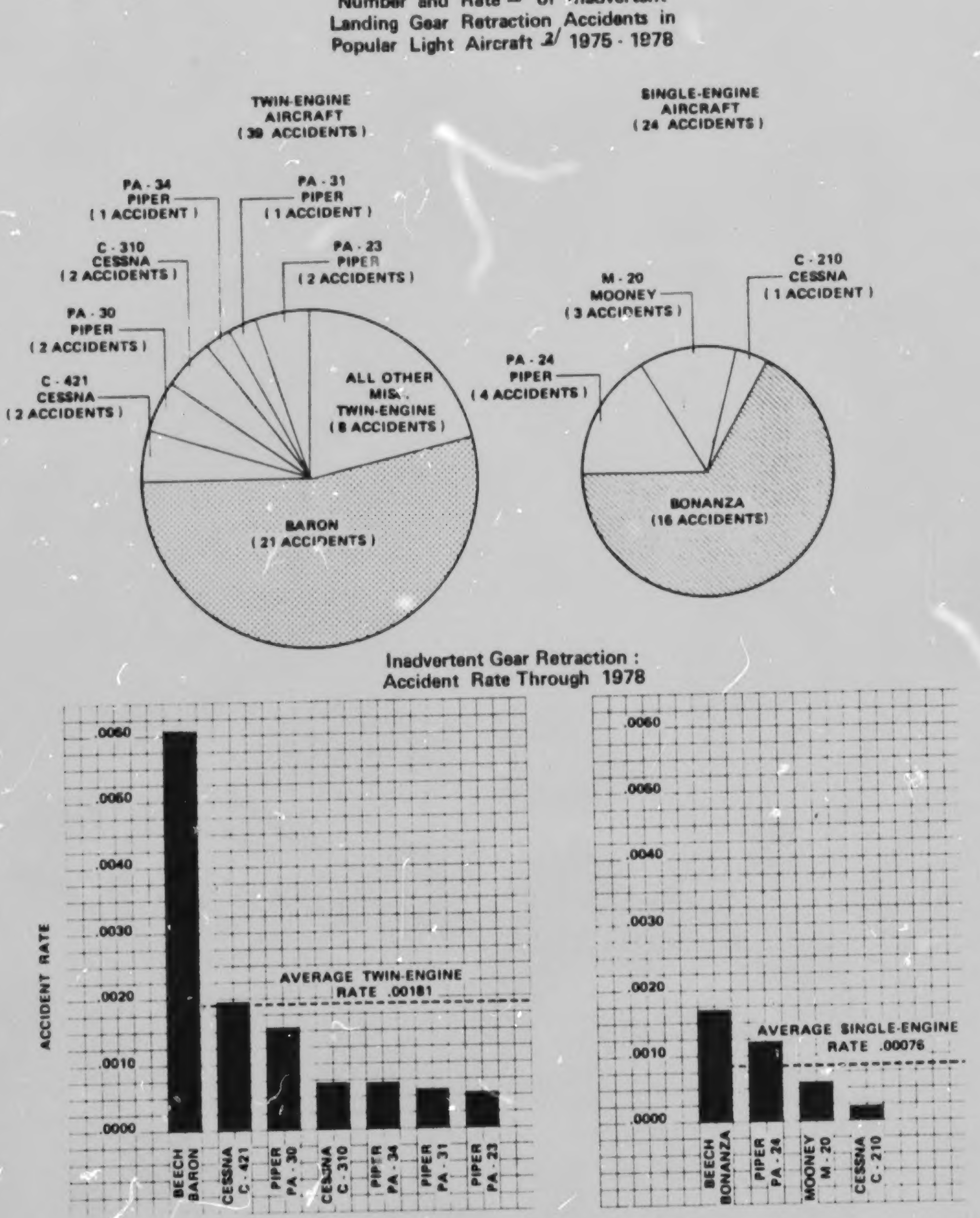
RETRACTABLE LANDING GEAR ACCIDENTS:
TWIN ENGINE AIRCRAFT

Table 2

Model	Date	Location	Pilot Total Hours in Accident Involved Model	Pilot Total Hours in All Makes and Models	Pilot Admitted Confusing Flaps With Landing Gear	Pilot Stated a Familiarity For a Reversed Arrangement of Gear and Flaps
BEECHCRAFT BARON						
BE-55	2/20/75	Kansas City, MO	47	1114	—	—
BE-58	3/5/75	Plymouth, MA	32	12586	X	X
BE-58	6/23/75	Phoenix, AZ	5	6580	X	—
BE-55	1/17/75	Anchorage, AL	99	7567	—	—
BE-58	8/20/75	Blountstown, FL	418	3308	—	—
BE-58	9/25/75	Jacksonville, FL	405	12220	X	—
BE-58	9/29/75	Little Rock, AR	12	872	X	—
BE-55	1/31/76	Fresno, CA	27	8300	—	—
BE-58P	7/20/76	Albuquerque, NM	15	12000	X	X
BE-95	1/3/77	Las Vegas, NV	40	700	X	X
BE-55	5/5/77	Davenport, IA	934	6841	X	—
BE-58	8/7/77	San Antonio, TX	12	1412	X	X
BE-58	11/2/77	Albany, NY	3	10660	X	—
BE-58	12/10/77	Loredo, TX	425	1400	X	—
BE-58	5/22/78	Kalakag, AL	18	14500	X	X
BE-58	5/31/78	Little Rock, AR	594	1205	X	—
BE-55	6/16/78	Walla Walla, WA	45	1232	X	X
BE-55	7/11/78	Albuquerque, NM	140	2300	X	—
BE-58	8/16/78	Hickory, NC	700	8355	X	X
BE-95	9/23/78	Amarillo, TX	100	6000	X	—
BE-55	12/24/78	Crosscut City, FL	1200	2200	X	—
MISCELLANEOUS TWIN ENGINE MODELS						
PA-23	6/3/75	Plattsburgh, NY	15	450	X	X
C-421	6/23/75	Chattanooga, TN	100	100	—	—
BE-50	2/11/76	Jacksonville, FL	7	1633	—	—
C-320	4/15/76	Granbury, TX	7	5000	X	X
C-310	7/21/76	New Samyra Beach, FL	25	904	X	—
AS-600	8/17/76	West Mifflin, PA	466	1592	—	—
C-421	9/12/76	International Falls, MN	74	2000	—	—
PA-23	9/16/76	Denopolis, AL	187	3123	—	—
SA-26	10/20/76	Claire, MI	70	8000	—	—
PA-30	4/6/77	Tuscaloosa, AL	450	2040	—	—
PA-30	4/9/77	Brooksville, FL	57	961	—	—
PA-40	6/22/77	Asheville, NC	240	12017	—	—
C-310	10/5/77	Cairo, GA	28	2587	—	—
PA-34	4/4/78	Cheyenne, WY	275	3628	—	—
E-4500	7/29/787	Togiac, AL	No Data		—	—
AC-500	7/31/78	Boston, MA	27	3762	—	—
BE(C-45)	8/8/78	Las Vegas, NV	144	9169	—	—
PA-31	10/9/78	Concord, NC	1555	8575	—	—

KEY

AC	- Aero Commander	C	- Cessna	SA	- Swearingen
AS	- Aerostar	E	- Evangle		
BE	- Beech	PA	- Piper		



1/ These rates were derived for each type aircraft by dividing the number of inadvertent landing gear retraction accidents by the estimated number of those aircraft which were active.

2/ Aircraft models with more than 500 active aircraft in 1978.

Figure 1.—Number and Rate of Inadvertent Landing Gear Retraction Accidents in Popular Light Aircraft, 1975-1978.

16 percent of the 1978 light-twin fleet, but suffered only 8 percent (2 accidents) of such mishaps occurring to light twins from 1975 to 1978. In contrast, the Baron, also comprising 16 percent of the twin fleet, experienced 67 percent of such mishaps (16 accidents) -- an accident rate of about 8 times that of the PA-23. Using the same differences in proportions test produced a "z" equal to 4.1 thus the Baron accident rate was significantly worse at the .01 level of confidence. Using the same standard statistical test again to compare the population parameters, it can be shown that the Bonanza accident rate for the such mishaps is significantly higher than the average for all single-engine aircraft at the .01 level of confidence, (z=2.6). Similarly this test indicates that the Baron accident rate is statistically worse than that of the average light twins at the .01 level of confidence (z=4.8).

A review of accident files for the 63 accidents occurred from 1975 to 1978 revealed several other facts. Tables 1 and 2 indicate that there appeared to be little correlation between pilot experience, either in total hours or hours in type, and the occurrence of these accidents. The data from Tables 1 and 2 also indicated that in 63 percent of the Bonanza accidents and in 81 percent of the Baron accidents, the pilots specifically admitted that they confused the landing gear and flaps controls. In many cases, they mistakenly retracted the gear while intending to raise the flaps after landing. Such explanations were only offered by 15 percent of the pilots following accidents in the other type aircraft.

An analysis of the individual NTSB case files revealed various circumstances which may have contributed to many of these accidents. Some pilots were either in stressful situations (such as danger of running off the runways) or they were distracted (such as by a tower controller's request to clear the active runway), or they may have been inattentive (such as when returning from a fatiguing flight).

HUMAN FACTORS ENGINEERING CONSIDERATIONS

There are numerous documents which describe the use of human engineering design features to decrease design-induced pilot error accidents. For example, a classic study, (Fitts and Jones, 1947) which surveyed hundreds of military pilots, found that confusing the flaps and landing gear controls was the second most frequent type of pilot-error control problem. The previously noted study, titled "Aircraft Design-Induced Pilot Error," (NTSB, 1967) was a comprehensive document detailing many of these problems, including the increased number of inadvertent gear retraction accidents resulting from certain aircraft design features.

The accidents reviewed in the paper illustrate the need for rigid adherence to procedures, constant vigilance, and total familiarity with the cockpit layout on the part of the pilot. However, they also underscore how design deficiencies can add to a pilot's burden and increase the likelihood of an accident. The following pilot statements extracted from Safety Board accident files, illustrate these points.

Bonanza, Elko, Nev., January 19, 1975: "When I reached to retract the flaps, I hit the gear switch instead. I also own a PA-30 in which the switches are in reverse to the Beech."

Baron, Plymouth, Mass., March 5, 1975: "I have thousands of hours in aircraft in which the flap switch is located where the gear switch is on the B-58 which was a contributing factor."

Baron, Las Vegas, Nev., January 1, 1977: "During rollout, at about 35/40 kts, pilot (me) retracted gear thinking it was the flap switch. Pilot used to flying Cessna 210 and flap switch is located where gear switch is located on Baron. Dumb pilot error."

Baron, San Antonio, Texas, August 7, 1977: "More careful familiarization with the instrument panel set up. This aircraft had a reverse set up for flaps and gear handles than the operator was used to."

Baron, Hickory, N.C., August 16, 1978: "Reached to retract flaps as for short field procedures, however, flap switch on Baron is reversed with landing switch on Cessna and Queen Air, pilot retracted landing gear instead of flaps."

Cessna 320, Granbury, Texas, April 4, 1976: "I have been flying a Bonanza and the gear and flap switch positions on Bonanza are exactly opposite to Cessna 320. . .. Require all manufacturers to place important controls consistently. Can you imagine a Cadillac and a Lincoln with brake and throttle in opposite positions?"

The regulatory requirements for the location and shape coding of these controls were first adopted October 1, 1959, by Amendment 3-5 to the Civil Air Regulations , which revised Section 3.384. These regulations were essentially identical to the current Federal Aviation Regulations (FARs) adopted in September 28, 1964, which require that the location and shape-coding of controls be standardized as follows: FAR 23.777 states: "Wing flap and auxiliary lift device controls must be located -- (1) Centrally, or to the right of the pedestal or powerplant throttle control centerline; and (2) Far enough away from the landing gear control to avoid confusion." The landing gear control gear must be located to the left of the throttle centerline or pedestal centerline. FAR 23.781 requires that cockpit controls must conform to the general shapes of wheel and airfoil for the landing gear and flap controls respectively.

The Bonanza was first type-certificated in 1945 and later recertificated in 1956. Also in 1956, the nonpressurized Barons were first type certificated. At that time, the Civil Air Regulations did not specify location or shape of the landing gear and flap controls. In 1959, the regulations were amended but the Bonanza and nonpressurized Barons were not required to meet the amended regulations and therefore continued to be produced under the earlier type certificates. The pressurized Barons were certificated in 1974 under FAR part 23, and therefore had to meet the requirements for the location and shape of these controls.

The design features of various types of aircraft involved in inadvertent landing gear retraction accidents were reviewed during this study. The importance of good human engineering in reducing the probability of pilot error has been noted previously (e.g. NTSB, 1967, Diehl, 1971, Ontiveros, Spangler and Sulzer, 1977).

An examination of cockpits of the Bonanza and Baron revealed four particular design deficiencies with regard to their landing gear and flap controls which can lead to design-induced pilot errors. These deficiencies include: (1) A lack of adequate "shape-coding" of these control knobs to permit the pilot to differentiate between them on the basis of feel alone; (2) an arrangement of these two controls in nonstandard locations which increases the probability that the pilot will actuate one control while intending to actuate the other; (3) the location of the horizontal bar on which the control wheels are mounted so that it obscures the pilot's view and obstructs his reach of these two controls; and (4) the lack of a guard or latch mechanism over the landing gear control to prevent the pilot from activating this control unless the guard/latch is moved first.

While various other types of modern light aircraft may have one of these four problems, the Bonanzas and Barons are the only aircraft produced in recent years with multiple combinations of these design deficiencies. (See Table 3.) The significance of the four type of deficiences should be noted:

Table 3.

Design Deficiencies for Different
Bonanza and Baron Models

Design Deficiency	Bonanza (pre-1963)	Bonanza (post-1963)	Baron (Nonpressurized)	Baron (pressurized)
Inadequate Shape-Coding	X			
Nonstandard Location	X	X	X	
Obscuration of Controls	X	X	X	X
Lack of Guard Latch		X	X	X

Inadequate Shape-Coding: The significance of shape-coding to reducing pilot error was clearly recognized in Fitts and Jones, (1947) which recommended shape-coding to prevent such errors. Classic research studies in Jenkins, (1947) have shown: (1) How certain knob shapes can be distinguished solely on the basis of touch, and (2) how by using symbolic shape associations which are similar to the function of the control (i.e. wheel-shaped knob for landing gear) the probability of misuse can be minimized.

The lack of shape-coded control knobs has been documented on the early Bonanzas in a study cited previously (NTSB, 1967). In describing these switches this report stated (p. 69) that ". . . the landing gear control and wing flap control are included in a row of similar switches or more precisely, nearly identical switches." The accident rate of the Bonanza was more than twice the average rate for all aircraft with retractable landing gears. When Beech redesigned the Bonanza cockpit in 1963, they did incorporate full shape-coding on these controls, but they deleted the latch which had been incorporated on previous models.

Nonstandardized Control Location: The significance of standardized locations to reducing pilot error was also clearly described in Fitts and Jones (1947). As with shape-coding, this document recommended standardizing the location of these controls to prevent errors. A recent FAA study (Ontiveros, Spangler and Sulzer, 1977) states (p. 1-E) that " . . . increased standardization of cockpit systems can reduce cockpit workload, reduce the potential for habit interference when transitioning to another type aircraft, and provide for application of the best and most error-resistant designs."

The detrimental effects of a nonstandardized control arrangement are illustrated by the contrasting accident rates of the Bonanza and the Cessna 210, which has a standard control arrangement. As shown by statistics, the Bonanza's inadvertent landing gear retraction accident rate is 10 times higher than that of the Cessna 210.

Obscuration of Controls: The problem of inadvertent gear retraction on the Bonanza and Baron aircraft is compounded further by a design feature of the flight control system which is unique to these two aircraft. The system utilizes a large horizontol cross-bar on which the control wheel (or wheels) is mounted. The two versions of this control system are (1) the single control wheel with a "throw-over" mechanism which allows the wheel to be placed in front of either the left or the right front seat, and (2) the dual control model where wheels are available to both seats.

There are two problems associated with this control system: (1) the horizontal bar is large enough to block the pilot's view of the gear and flap control switches, forcing the pilot to rely on his sense of feel to identify the desired control, and (2) the pilot must reach around the bar to activate these controls. Both of these problems are more of a hindrance to pilots of small stature and when the wheel is relatively far forward. The control switches are relatively small in comparison to those on many other aircraft. This also tends to decrease the pilot's ability to differentiate those controls by feel.

The presurrized Baron (58P), which was certificated in 1974 and meets FAR part 23 requirements with respect to landing gear and flap control location and shape-coding, was involved in only one landing gear retraction accident during the 1975 to 1978 period. Ironically, the pilot attributed his mistake in part to the fact that he was more familiar with the nonstandard control arrangement of the unpressurized Baron and Bonanza. However, he also pointed out that his view of these controls was blocked by the wheel-mounting mechanism.

Lack of a Landing Gear Control Guard Latch: The advantages of incorporating a latch or guard on the landing gear control can be seen by comparing the accident rate of the Baron with that of a similar aircraft, the Piper PA-23 Aztec. The PA-23 is the only other light plane currently being produced with a nonstandard gear and flap control arrangement. However, the landing gear control on this aircraft is protected from inadvertent actuation by a separate mechanical guard latch, and as noted earlier, its inadvertent landing gear retraction accident rate is only one-tenth that of the Baron.

CONCLUSIONS

The NTSB concluded that the number of inadvertent landing gear retraction accidents in the Beech Bonanza and Baron was unnecessarily high in comparison to other contemporary general aviation aircraft. The Board also concluded on the basis of various pilot statements, a review of the human factors research literature, and a detailed analysis of the cockpit features of these aircraft that these accidents result largely from various combinations of four design deficiencies - inadequate shape-coding, nonstandard location of controls, obscuration of controls, and lack of a guard latch on the landing gear control. There appears to be a growing recognition of the importance of human engineering in preventing accidents, by the public, various government agencies and aircraft industry. However, the occurrence of such accidents also illustrates the importance of the pilots' responsibilities to familiarize themselves with cockpit details, particularly when flying dissimilar aircraft.

REFERENCES

"Aircraft Design - Induced Pilot Error", National Transportation Safety Board, Special Study PB 175629, July, 1967.

Diehl, A.E., "Human Engineering and the Cost-Effectiveness of Air-Safety Devices." Society of Automotive Engineers, Publication Number 710487, March, 1971.

Fitts, P.E. and Jones, R.E. "Analysis of Factors Contributing to 460 'Pilot-Error' Experiences in Operating Aircraft Controls" USAF Aeromedical Laboratory, Memorandum Report, July, 1947.

Jenkins, W.O. "Tactile Discrimination of Shapes for Coding Aircraft - Type Controls" U.S. Army, Air Force, Aviation Psychology Program Research Report 19, 1947.

Ontiveros, R.J. Spangler, R.M. and Sulzer, R.L. "General Aviation (FAR 23) Cockpit Standardization Analysis", Report No. RD-77-192, March 1978.

Spiegel, M.R."Theory of Problems of Statistics,"McGraw-Hill, Inc., New York, New York, 1961.

IN-TRAIL FOLLOWING DURING PROFILE DESCENTS WITH A COCKPIT DISPLAY OF TRAFFIC INFORMATION

Sheryl L. Chappell*
Tufts University
Medford, Mass.

Everett A. Palmer
NASA Ames Research Center
Moffett Field, Calif.

Abstract

How well can a pilot maintain 100 seconds separation behind another aircraft using information from a cockpit display of traffic information (CDTI)? Four line pilots flew a B-747 simulator from cruise to landing at Denver, following five different aircraft profiles. The initial spacing of the aircraft varied from 60 to 140 seconds.

When a CDTI was used for in-trail following, initial spacing error was resolved fairly quickly. Spacing error, however, became excessive in the final stages of the approach, especially when the lead aircraft deviated from the nominal profile. Fuel consumption was no different with or without an aircraft to follow. Subjective ratings showed workload was higher while following another aircraft.

Introduction

A cockpit display of traffic information (CDTI) is a display in the cockpit showing the pilot the position of other aircraft and may include other information about the aircraft, e.g. ground speed and altitude. One question that arises with the advent of CDTI's concerns the changes possible in the responsibilities of pilots and controllers. This study examines the ability of a pilot to use the information from a CDTI to follow another aircraft and maintain a safe distance behind that aircraft. The in-trail following clearance given to the pilots is similar to a presently used type of visual approach clearance. The aircraft is cleared to follow another aircraft which is in sight and the pilot must maintain visual separation. Therefore there is a shift in responsibility for spacing from air traffic control (ATC) to the cockpit.

Methodology

Subjects

Four line pilots with major U.S. air carriers, based in the San Francisco area participated in this study. They are currently

---

* Supported by NASA Grant NSG 2156 to Tufts University.

assigned to the Boeing 747 as captain or first officer. The pilot serving as first officer was part of the experiment team. The experimenter performed the duties of the flight engineer. The air traffic controller is a pilot working on the project.

Equipment

The simulator is based on the Boeing 747 with all the necessary controls and flight, navigational and engine instruments. The pilots flew the simulator using the heading-select mode of the autopilot, and the pitch wheel or altitude select/hold mode. The throttle was always manual. All the landings were performed by the autopilot. The flight director was available as was raw data of VOR/localizer and glide slope. An intercom was used for the radio transmissions to the controller. No wind or turbulence was simulated.

The cockpit display of traffic information used in this study is presented on a CRT in front of the throttles. The display is always heading-up and depicts a map of the navaids and airways in the Denver area (Figure 1). The VOR radial that has been selected is shown with dotted lines in the center and 3 nautical miles on each side of the centerline, showing the approximate width of the airway. The range or map scale is twenty nautical miles from the ownship's aircraft symbol to the top, at altitudes above 18,000 feet, and ten nautical miles below that altitude. The airway appears to pass under the stationary ownship symbol.

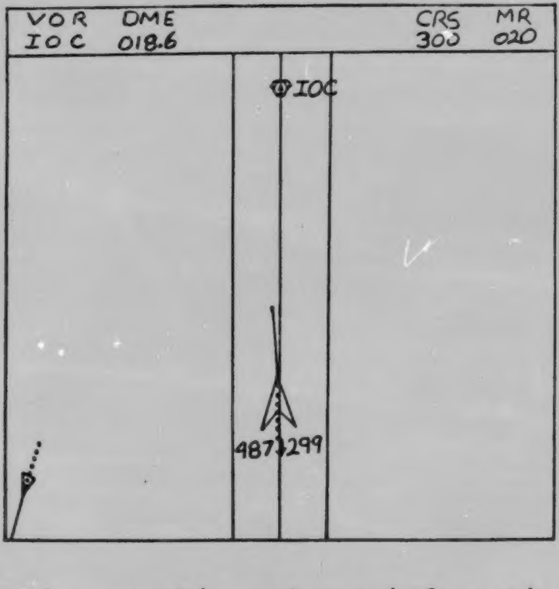


Figure 1. CDTI showing ownship and map information.

The apex of the aircraft symbol shows the actual position of the aircraft. The flight-path predictor line in front of the aircraft symbol shows where the aircraft will be in 60 seconds if present speed, heading, and rate of turn are maintained for that period of time. The curve of the predictor reflects the rate of turn. The length will vary with ground speed since it reflects the distance traveled in 60 seconds. The dots behind ownship show the previous positions at four-second intervals for the last forty seconds. The data tag shows ground speed (in knots) on the left, and altitude (in hundreds of feet) on the right. This configuration was chosen because in the cockpit, the airspeed indicator is to the left of the altimeter. In the center of these two numbers is an arrow which indicates whether the aircraft is climbing or descending, or when level (vertical velocity less than one meter per second) a dash will appear. All the information in the display discussed thus far, comes from instruments in the aircraft, e.g. turn and bank instruments, altimeter and navigation instruments. This information on the display is updated every tenth of a second.

The traffic is shown on this CDTI by a triangle pointing along the aircraft's ground track (Figure 2). The other aircraft also have a 60 second predictor, but it is always straight, even when they are turning. There is also a trail of dots showing where the aircraft was every 4 seconds for the last 40 seconds. The data tags for these aircraft are the same as the ownship. All the information on the traffic is updated every 4 seconds.

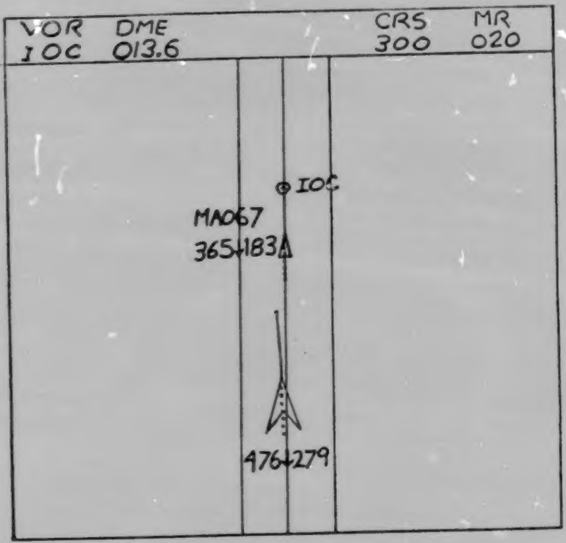


Figure 2. CDTI showing ownship and lead aircraft.

When this display is actually implemented the information about the other aircraft may be provided by ground radar and sent to the aircraft via data-link. For this study, no radar errors were simulated, but an alpha-beta tracker was used to estimate the velocity necessary for the traffic predictors.

Procedure

The pilots were paid for the four mornings they reported. There were six flights each day. The first day was for training, and the subsequent days, experimental. All the simulated flights were profile descents into Denver, Stapleton International Airport. Each flight began at cruise and consisted of a standard profile descent, ILS approach, and landing. The pilot's role was that of captain doing the flying, with a first officer and second officer complying with standard airline procedures. In addition, all the normal radio communications were carried out with the controller.

The flight began 97 DME (nautical miles) to Kiowa VOR and the aircraft was in contact with Denver Center at cruise airspeed and altitude. Figure 3 is the low profile descent chart for Denver*, with the crossing altitudes and speed restrictions at both Kiowa and Wifes intersection. After Wifes aircraft are vectored for the ILS to runway 26 left and fly the ILS according to the published procedures.

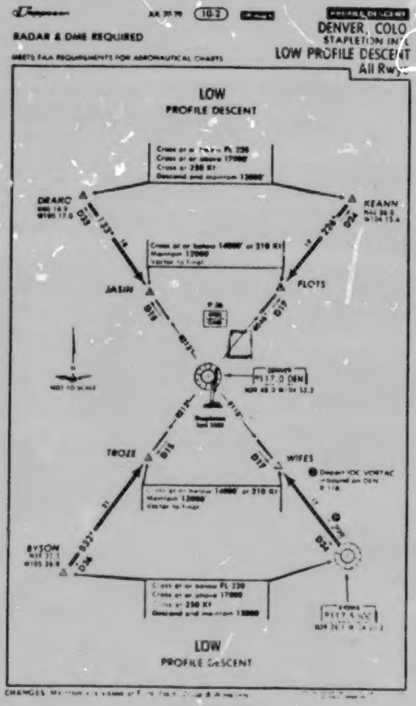

10-2
RADAR & DME REQUIRED
DENVER, COLO
STAPLETON INT'L
LOW PROFILE DESCENT
All Rwys
LOW
PROFILE DESCENT
Cross at or above 17000'
Cross at 250 Kt
Descend and maintain 13000'
DRAKO
KEANN
133°
226°
Cross at or below 14000' at 210 Kt
Maintain 13000
Vector to final
JASIN
FLOTS
117.0 DEN
NOT TO SCALE
TROZE
WIFES
BYSON
033°
Cross at or below FL 230
Cross at or above 17000
Cross at 250 Kt
Descend and maintain 13000
LOW
PROFILE DESCENT

Figure 3. Low Profile Descent Chart for Denver

---

* Low Profile Descent Chart, Denver Colo., Plate No. 10-2, Dated 7/27/79. Jeppesen Sanderson Inc., Denver (1978).

Pilots were permitted to fly the training flights manually but were asked to use the autopilot for the experimental flights. The autopilot was used (autothrottle not available) because during actual air-carrier operations much of the flight is made using the autopilot. The pilots were provided with minimum maneuvering speeds for the different flap configurations and the approach reference speed, calculated for the gross weight of the simulated aircraft.

Training. The first two flights were flown with no CDTI to familiarize the pilot with the simulator. The third flight was made with the CDTI and no other traffic displayed. On this flight, the pilots were not required to use the information on the display, but could refer to it for navigation as workload permitted.

On the fourth flight the traffic was introduced. On this flight the aircraft was far enough ahead so as not to require any adjustment of spacing. The pilots completed a questionnaire to assess their understanding of the information on the display and any misunderstandings were discussed.

For the fifth and sixth flights the in-trail following clearance was in effect. The rules regarding the following clearance were presented. These are based on the visual approach clearance to follow another aircraft. First, ATC confirms "electronic contact", i.e. the pilot has the other aircraft on the CDTI. Then the clearance is given for the pilot to follow the other aircraft at a given number of seconds. The pilot is expected to maintain the spacing within +/- 16 seconds (four dots of history) of the specified clearance. The altitude restrictions depicted on the published approach charts must be complied with, however, speed restrictions are not in effect when a spacing clearance has been given. While the aircraft is on radar vectors, ATC will provide headings and altitudes but the pilot is responsible for separation behind the aircraft being followed. And finally, as with any clearance, when unable to meet or maintain the required spacing, the pilot must notify ATC immediately and request further instructions.

Initially, the aircraft to be followed was either 60, 80, 120 or 140 seconds ahead. The clearance stated the pilot was to follow the identified aircraft at 100 seconds. This spacing remained in effect until the aircraft was cleared to land. When the tip of the 60-second predictor of the ownship is placed on the final history dot (40 seconds) of the lead aircraft the separation of the two aircraft is approximately 100 seconds.

Time was chosen as the dimension for spacing in order to vary the distance with ground speed. For example, a 100-second spacing at an approach speed of 140 knots (ground speed) results in 4 nautical miles separation and 14 miles at a cruise ground speed of 500 knots. This method provides greater distance separation at faster ground speeds.

In addition to the usual pilot-not-flying "call outs", it was suggested that the pilots work out some call-outs with the copilot regarding the status of spacing, e.g. "8 seconds too close".

Experimental Flights. There were six (randomly ordered) flights on each of the experimental days. One flight was made with no lead aircraft but the CDTI was available for navigation. One flight began 100 seconds behind the lead aircraft and the lead aircraft flew a nominal profile descent and ILS approach, i.e. totally complied with the published procedures and started the idle descent so as to meet the crossing restrictions with minimal use of power.

The other four flights had an initial error in the spacing. Also, the lead aircraft made slight miscalculations in the descent resulting in a profile that deviated some what from the nominal. Two flights began at 80 or 120 seconds (100 +/- 20) behind the preceding aircraft. The lead aircraft either began the descent too soon and leveled off and slowed down before Kiowa VOR, or began descent later than the optimum time. The late descent resulted in a high vertical velocity and ground speed at Kiowa VOR. The two other flights were the most difficult for maintaining 100 seconds separation. These flights began at 60 or 140 seconds (100 +/- 40) initial spacing. In these flights the lead aircraft flew a nominal descent profile to Kiowa, then either slowed to 160 knots (ground speed) at Wifes intersection (17 DME to DEN), or was 1000 feet high and too fast at the outer marker.

At the end of each flight, time was taken for discussion and questions. Also, a set of workload evaluation scales was completed by the pilot. The pilot was asked to evaluate the workload for the preceeding flight on a seven-point scale. The contrasting adjectives used were:

Demanding	_ _ _ _ _ _ _	Undemanding
Subtle	_ _ _ _ _ _ _	Obvious
Unpredictable	_ _ _ _ _ _ _	Predictable
No Skill Required	_ _ _ _ _ _ _	Much Skill Required
Easy	_ _ _ _ _ _ _	Difficult
High Stress	_ _ _ _ _ _ _	Low Stress
Busy	_ _ _ _ _ _ _	Idle
Few Interruptions	_ _ _ _ _ _ _	Many Interruptions
No Uncertainty	_ _ _ _ _ _ _	Much Uncertainty
No Planning Needed	_ _ _ _ _ _ _	Much Planning Needed
High Risk	_ _ _ _ _ _ _	Low Risk
Satisfactory Performance	_ _ _ _ _ _ _	Unsatisfactory Performance

At the beginning of each experimental day, there was a review of the clearance procedures and the information on the CDTI.

Results

There are several areas of interest when exploring the use of a CDTI for following another aircraft during an approach. The foremost question is safety. Does the implementation of the in-trail following clearance increase or decrease the overall safety of air travel? More specifically, is the pilot able to perform the task within a safe tolerance? If so, is there an increase in workload? Is any workload increase acceptable? Is there sufficient redundancy of information to allow error detection without jeopardizing safety? Among other concerns is economics, e.g. how is the rate of fuel consumption effected by the in-trail following clearance?

Performance Variables

There were two performance variables used to evaluate the ability of the pilot to perform the in-trail following task: spacing error and fuel consumption.

Spacing error is any deviation from the 100 seconds separation required by the approach clearance. The difference was measured between the two aircraft's arrival times at a position. Spacing error is that time minus the 100 seconds desired separation. This error was calculated at 16 positions along the course of the descent and approach. The amount of spacing error has importance for safety and efficiency of traffic flow. If the pilot flies too close to the preceding aircraft, there is a danger of collision. If the aircraft are too far apart in spacing, fewer aircraft can land in the same amount of time. Therefore, there is a decrease in the capacity of the air traffic system.

Figures 4-6 show the spacing error throughout the flights for the different conditions of spacing error at the start of the flight and the different lead aircraft that were followed in trail. These data show that the initial errors in spacing are reduced in about 60 miles after which the error remains small (and generally positive) for about 50 miles. As the flight continues beyond Wifes intersection (17 DME), the amount and the variance of the error greatly increase.

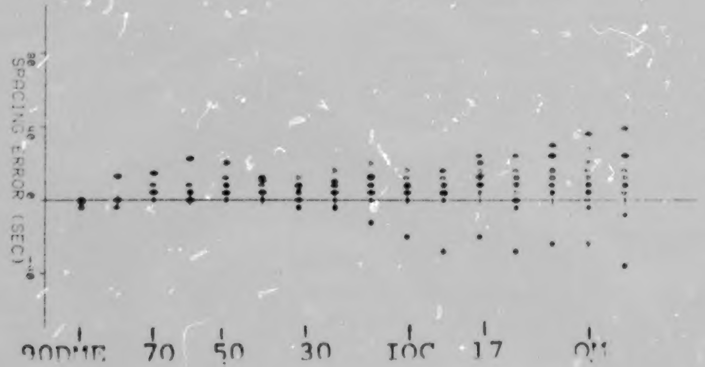


Figure 4. Spacing error for flights with 100 second initial spacing error behind Lead aircraft 1.

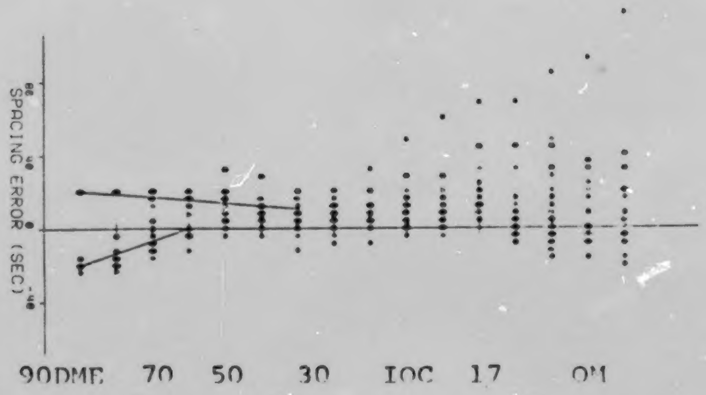


Figure 5. Spacing error for flights with 80 & 120 second initial spacing error behind Lead aircraft 2 & 3.

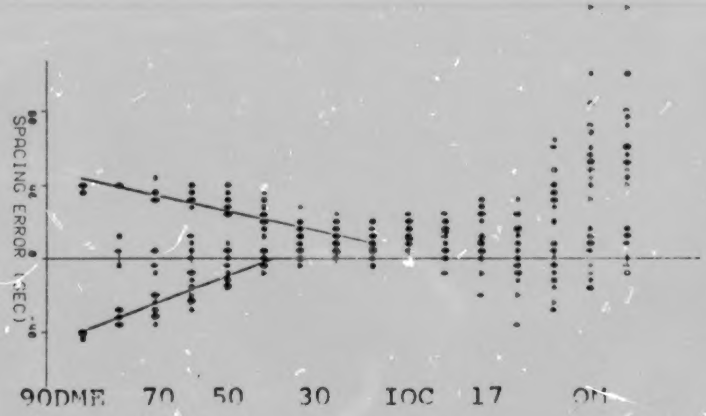


Figure 6. Spacing error for flights with 60 & 140 second initial spacing error behind Lead aircraft 4 & 5.

The first step in the analysis of the spacing error data was to determine if the initial spacing condition and the lead aircraft that was followed were independent in their effects on spacing error. These effects were purposely confounded in the experimental design. A t test on the spacing error mean (of the 16 positions) for each flight, found no difference due to initial spacing with any lead aircraft (p > .1). Conversely, there was no difference due to lead aircraft with any initial spacing (p > .1). The initial spacing effect was resolved early in the flight (after approximately 60 miles), whereas the lead aircraft only differed in effect on spacing after about 90 miles.

Initial Spacing Condition. One question that arises when evaluating pilots' ability to perform the aircraft-following task is: how long does the pilot take to resolve a discrepancy in separation? For the flights which began with an error in spacing, the number of miles was determined for that error to be reduced to a minimum. Initial spacing error was resolved more quickly when the aircraft were too close to one another. Due to the aircraft performance characteristics at 35,000 feet, the aircraft can slow down or make "S" turns more readily than it can increase its cruise speed. Figures 4-6 show the spacing error for the different initial spacing conditions 100 seconds (0 seconds error), 80 & 120 seconds (-20 & +20 seconds), 60 & 140 seconds (-40 & +40 seconds). The regression lines drawn are for the spacing error from the beginning of the flight to the point of minimum spacing error. The spacing error taking the longest to resolve was the +40 second initial condition. This error was minimized in 80 miles. The most quickly reduced initial spacing error was 20 seconds too close. This error was minimized in 30 miles. Errors of +20 and -40 seconds were minimized in 60 miles.

Discrepancies in the 100-second separation were resolved at a similar rate whether the initial error was 40 seconds or 20 seconds. This may be seen in the similarity in the slope of the regression lines of the spacing error when the aircraft started 20 seconds too close (Figure 4) or 40 seconds too close (Figures 5) to one another. The rate of spacing error reduction was approximately similar when the aircraft started 20 and 40 seconds too far apart.

The spacing error (absolute value) was compared for the first half of the flight (80 miles). All pairs of initial spacing conditions were significantly different in spacing error (t test $p < .001$) except the +20 seconds and -40 seconds conditions. These two initial spacing conditions were minimized at the same point in the flight.

Lead Aircraft Condition. The lead aircraft's deviation from a nominal profile caused a spacing perturbation during the last half of the flight (Figures 7-11). Lead 1 was a nominal profile. Lead 2 and 3 began descent too early and late, respectively. Lead 4 slowed to 160 knots at Wifes intersection. Lead 5 was too high and fast at the outer marker.

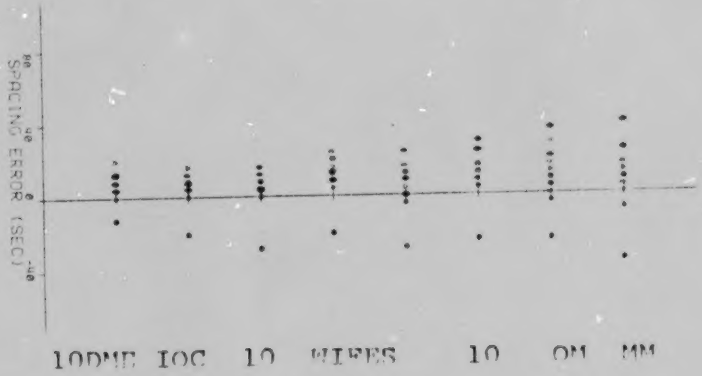


Figure 7. Spacing error during the second half of the flights (from 10 DME before Kiowa) behind Lead aircraft 1 (nominal profile).

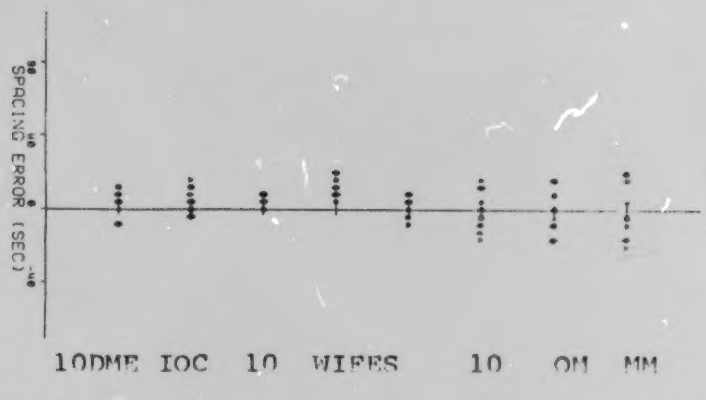


Figure 8. Spacing error during the second half of the flights (from 10 DME before Kiowa) behind Lead aircraft 2 (began descent too early).

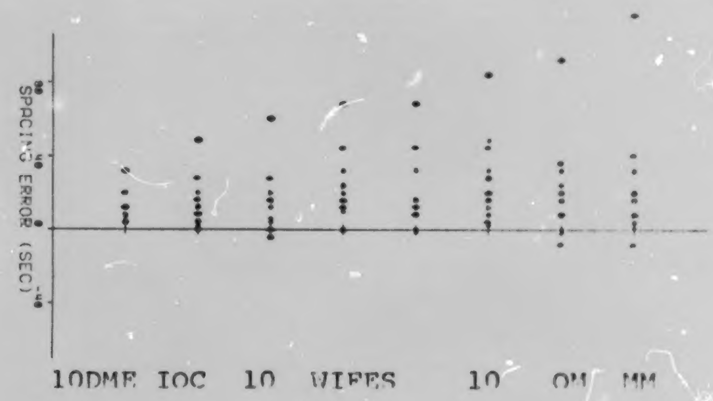


Figure 9. Spacing error during the second half of the flights (from 10 DME before Kiowa) behind Lead aircraft 3 (began descent too late).

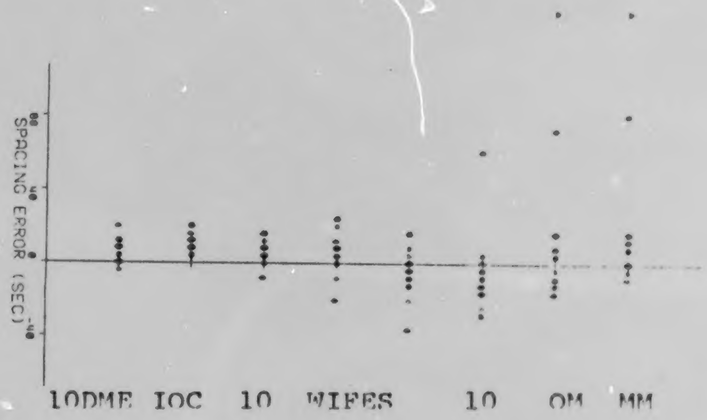


Figure 10. Spacing error during the second half of the flights (from 10 DME before Kiowa) behind Lead aircraft 4 (slow at Wifes).

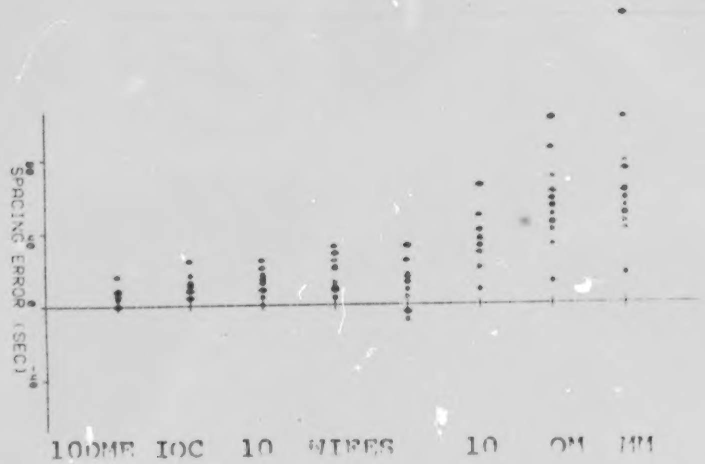


Figure 11. Spacing error during the second half of the flights (from 10 DME before Kiowa) behind Lead aircraft 5 (high and fast at the outer marker).

A comparison was made of the absolute value of the spacing error for the second half of the flight between all pairs of lead aircraft conditions. Only two pairs of lead aircraft were not significantly different (T statistic $p < .05$). These were Lead 4 with Lead 1 and with Lead 2.

Since the lead aircraft profiles are all variations that could be expected on an approach, the level and variance in spacing error is a critical consideration. Figures 7-11 show the spacing error for the last half of each flight following different lead aircraft. Pilots flew their aircraft too close to Leads 2 and 4, thus reducing the margin of safety. Separation behind Leads 3 and 5 became excessively large after Wifes. This has little impact on safety but decreases the efficiency of traffic flow. Lead aircraft 1 on the nominal profile produced a gradual increase in positive spacing error with no large changes in slope.

Fuel Consumption. A major concern in today's air carrier operations is the amount of fuel consumed during a flight. Fuel consumption during a descent is also an indication of the amount of throttle and speed adjustments. In a well planned descent, idle thrust can be used until established on final approach, unless traffic requires speed or course adjustment. A comparison between the fuel used on flights with no aircraft to follow and flights with different initial spacing conditions yielded no difference (t tests $p > .5$). Flights beginning 100 seconds behind the lead did not differ in fuel consumed with flights having an initial spacing error (t tests $p \geq .1$).

The fuel used on flights with no lead aircraft was then compared with the fuel used when following the five different lead aircraft. There was no difference in the amount of fuel consumed (t tests $p > .1$). Fuel consumption for flights following the nominal lead was no different (t tests $p > .1$) than flights following the other lead aircraft except for Lead aircraft 4 ($p < .05$, see Figure 10). Adjustment in spacing behind Lead 4 required a reduction in speed at an altitude of approximately 14,000 feet. Pilots had to lower flaps and gear earlier in order to reduce speed and increase separation, resulting in an increase in the use of power.

Subjective Workload Measures

Once the ability of a pilot to perform the task has been determined, an important question to consider is the level of the pilot's workload while performing that task. The workload rating scales that were completed after each flight showed significant difference between the level of workload for a flight using the CDTI to follow 100 seconds behind another aircraft and those flights where there was no other aircraft. Especially the adjective pair Demanding/Undemanding, which had a Wilcoxon Signed Ranks Test $p < .05$ for no lead aircraft compared to all lead aircraft conditions and all initial spacing conditions. Other adjective pairs also showed differences in workload ratings between flights with no other aircraft and the in-trail following flights. However there were only a few isolated differences among the different lead aircraft conditions and among the initial spacing conditions.

Learning Effects. The spacing error continued to decrease on each of the three experimental days (one-way analysis of variance $F=6.28$ $p < .01$). The mean spacing error reduced from 16.61 to 14.13 to 12.20 on the subsequent days. However, the subjective workload evaluations of flights made on each of the three days showed no difference (Kruskal-Wallis one-way analysis of variance $p > .1$).

Discussion

The findings suggest that pilots are able to follow another aircraft in trail using the information from a CDTI. They are able to reduce error in the initial aircraft spacing rather quickly, given the constraints of aircraft performance. Generally the separation error remained low until the point at which the pilot began making preparation for landing. The increase in error in the final stages of the approach suggests that the level of workload has a detrimental effect on the pilot's performance of the in-trail following task. The observations of the experimenters support this hypothesis. As the pilot became more involved in the control of the aircraft in preparation for landing, less time was spent looking at the CDTI. The workload evaluation scales do not specifically address different phases of the flight, however the pilots rated in-trail following flights in general as having higher workload than flights with no other aircraft.

In conclusion, the CDTI provides the necessary information to enable a pilot to maintain separation from the preceding aircraft. The pilot can take the responsibility for spacing rather than having the spacing accomplished by ATC. This information expands the capability of the pilot under instrument conditions to that similar in a visual environment.

PAVE LOW III
INTERIOR LIGHTING RECONFIGURATION FOR NIGHT LIGHTING AND NIGHT VISION GOGGLE COMPATIBILITY

H. L. TASK
L. L. GRIFFIN

Human Engineering Division
Aerospace Medical Research Laboratory
Wright-Patterson AFB, Ohio

ABSTRACT

The PAVE LOW III aircraft is a modified HH 53-H helicopter that has a low altitude (below 100 feet), night/day rescue mission. The desired night flying configuration is for the pilot to wear night vision goggles (NVGs) to fly the aircraft while the copilot, without NVGs, observes displays and monitors the aircraft instruments.

The problems of NVG incompatibility in the cockpit were successfully countered using several light control techniques.

The light control modifications were evaluated on the ground in the PAVE LOW III helicopter at Kirtland AFB in April 1980 by PAVE LOW instructor pilots. The evaluation results were extremely positive.

INTRODUCTION

The PAVE LOW III aircraft is a modified version of the HH-53H helicopter (see Figure 1). Its primary mission is day/night air rescue. The mission profile of this aircraft is to fly extremely low for day/night search and rescue of downed pilots. The original PAVE LOW III modifications included a forward-looking infra-red (FLIR) imaging sensor mounted on a moveable gimbal at the forward section of the aircraft. This FLIR provided night, infra-red imagery via two 5" x 7" cathod-ray tubes (CRTs) mounted in the instrument panel in

Figure 1. The PAVE LOW III Aircraft; a Modified HH-53H Helicopter.

front of the pilot and co-pilot (see Figure 2). Additionally, to support night and adverse weather navigation, a radar altimeter and Terrain Avoidance/Terrain Following radar was installed. However, as the PAVE LOW III aircraft was undergoing acceptance testing and required participation in Red Flag '79, (Tactical Air Combat Exercise), the requirement for low altitude flight was extended beyond the design capablities of the radar. It was felt by those familiar with the helicopter that lower altitudes could be achieved at night if the pilot used the U.S. Army developed, second generation Night Vision Goggles (NVGs). Figure 3 shows a pair of the AN/PVS-5 night vision goggles.

Figure 2. The Front Instrument Panel and Center Console of the PAVE LOW III.

In initial tests with the night vision goggles it was determined that the interior lighting for night flight interfered severely with

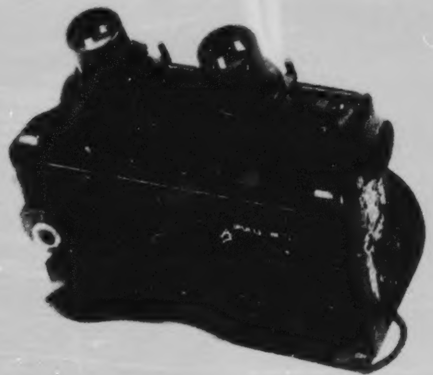


Figure 3. Army Developed AN/PVS-5 Second Generation Night Vision Goggles.

the useful operation of the NVGs. The night illumination, even adjusted to a low level, emits considerable energy in the near infrared where the NVGs are most sensitive. A first attempt to reduce this problem was conducted by the PRAM (Productibility, Reliability, Availability, Maintainability) PO (Program Office) at Wright-Patterson AFB in cooperation with the Military Airlift Command (MAC) and the Air Rescue and Recovery Service (ARRS). This initial test involved turning off all possible interior lights and floodlighting the instrument panel with yellow-green electro-luminescent lighting. Electro-luminescent (E-L) light emits almost all of its energy in the visible region and essentially none in the infra-red. This "cold light" effect makes the E-L light much more compatible with the use of NVGs than the traditional incandescent lighting.

At a meeting late in 1979 the authors were asked by PRAM and MAC to address the problem of making the interior lighting of the PAVE LOW III helicopter compatible with the use of the NVGs. The desired operating condition was for the pilot to wear the NVGs to fly the helicopter while observing the outside world, and for the copilot to monitor the FLIR video display and the aircraft instruments. Thus the fundamental problem was to design a means of lighting such that the copilot had sufficient light to monitor the cockpit instruments but insure that the lighting did not interfere with the pilot's NVGs. A review of the aircraft interior lighting and the windscreen/instrument geometry revealed two sources of lighting difficulty. First, several illuminated instruments on the center and overhead console reflected directly in the windscreen from the pilot's and copilot's eye position as well as the flight engineer's nominal eye position. This problem makes night flight, even without the NVGs, difficult and distracting and almost totally disallowed the use of NVGs. Second, the stray light in the cockpit illuminated other surfaces (like the pilot's knees and hands) such that their reflections in the windscreen were highly visible and distracting when viewing through the NVGs. To improve the NVG utility it was necessary to eliminate the direct reflecting sources and reduce or control the scattered light.

APPROACH

Several lighting and light control techniques were recommended to alleviate the NVG incompatibility problem:

1. Use blue-green E-L flood-lighting and turn off all possible incandescent lamp sources.

2. Use blue filters over the CRT display (instead of red) and place a red filter over the NVGs.

3. Use light baffles to control stray light.

4. Use anti-flare baffles on NVGs to reduce flare.

5. Use flat black flight clothing and helmets to reduce stray light and reflections.

As a result of the meeting with MAC and PRAM, the authors were requested to impliment the above recommendations on a PAVE LOW III helicopter for a full-up ground evaluation. The following sections describe each of these recommendations in detail.

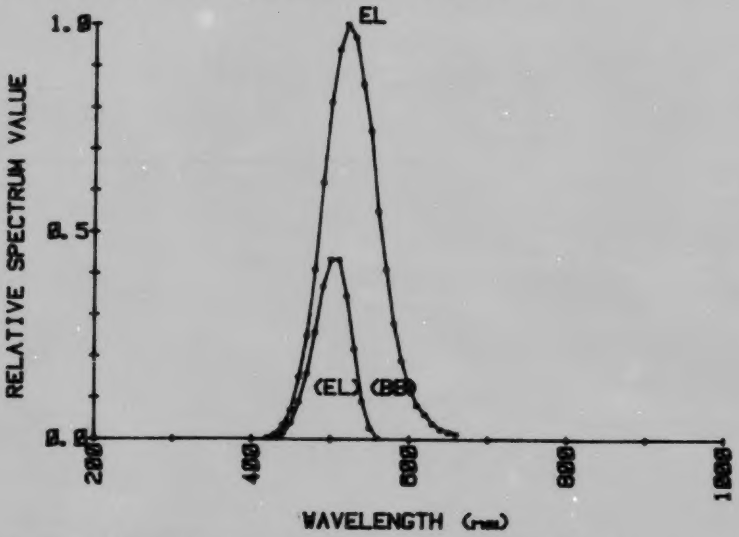


Figure 4. Wavelength Emission Spectrum of the Electro-Luminescent Lamp Without Filter (E-L) and With a Blue Filter (BB).

Electro-Luminescent Flood Lighting

The main reason for using E-L light is that, unlike incandescent lighting, it emits little or no light in the near infra-red region. The NVGs are highly sensitive to light from about 450 nanometers to 850 namometers wavelength. By limiting the interior lighting to blue-green, visible wavelengths only, considerable adverse interaction between the lighting system and the NVGs was eliminated. Figure 4 shows the emission spectrum of the E-L lamps used. The upper curve is the E-L lamp without filter and the lower curve shows the lamp with a blue filter (BB) that was used to shift the emission spectrum further into the blue region.

The E-L lamps were placed under the glare shield to illuminate the front instrument panel and center console. They were also placed on the backs of the pilot's and copilot's seats, directed upwards, to illuminate the overhead panels. It was not possible under the constraints placed on this retrofit (no holes drilled or permanent modifications allowed) to properly illuminate the far forward section of the overhead panels or the rear section of the center console. Thus to provide a means of "portable" illumination, an E-L light wand was provided that the copilot could use for map reading or close-up instrument tasks such as setting radio frequencies. In an ideal situation these areas would be locally illuminated with E-L light built into the instrument or its immediate surround.

Blue/Red Filters

The CRT FLIR displays on the helicopter use a P-4 white phosphor. Although this emits no infra-red light it does emit over the full visible region. The standard night operation required the white CRT screen to be covered with a red filter. This left a display emission spectrum in the visible from about 600 nm to 650 nm. This spectrum is in the center of the sensitivity region of the NVGs. Figure 5 shows

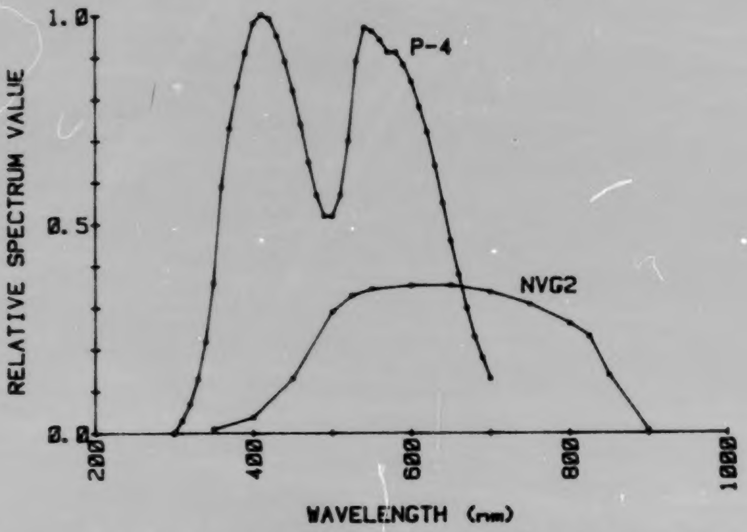


Figure 5. Composition of the Display P-4 Phosphor Spectrum and the Night Vision Goggle (NVG 2) Sensitivity Curves.

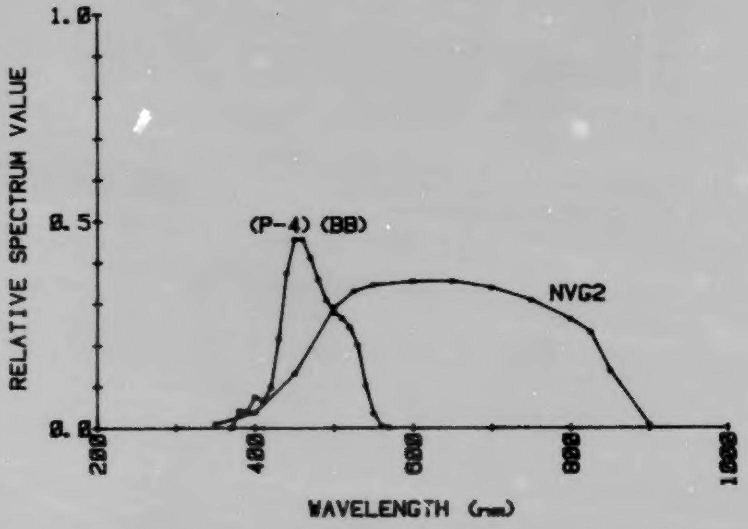


Figure 6. Comparison of P-4 Phosphor with Blue Filter and Night Vision Goggle Sensitivity Curves.

the relative sensitivity of the second generation night visions goggles (NVG2) compared with the emission spectrum of the P-4 phosphor.

By using a blue filter (BB) over the P-4 phosphor screen it is possible to shift the emission spectrum toward the blue, where the NVG is not quite as sensitive (see Figure 6). This still results in considerable overlap. To reduce the overlap still further, a red plastic filter was placed over the NVGs that was also highly transmissive in the near infra-red. This resulted in the curves shown in Figure 7. Under these conditions the emission of the display and the

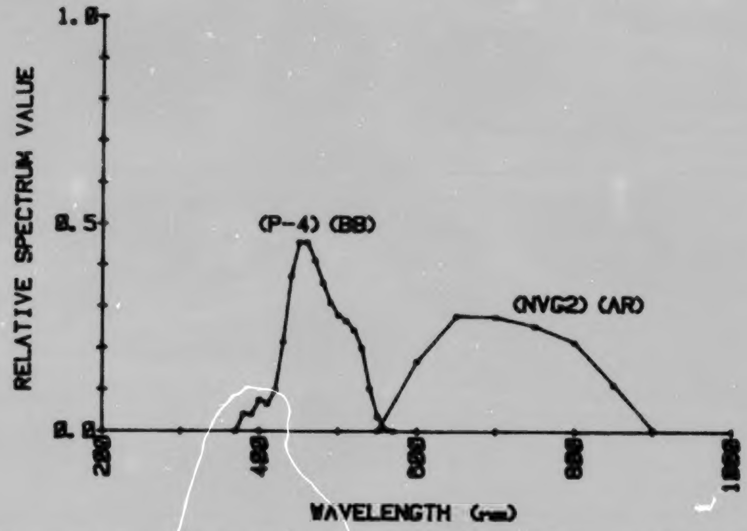


Figure 7. Comparison of P-4 Phosphor with Blue Filter and Night Vision Goggle Sensitivity with Red Filter. Note lack of overlap.

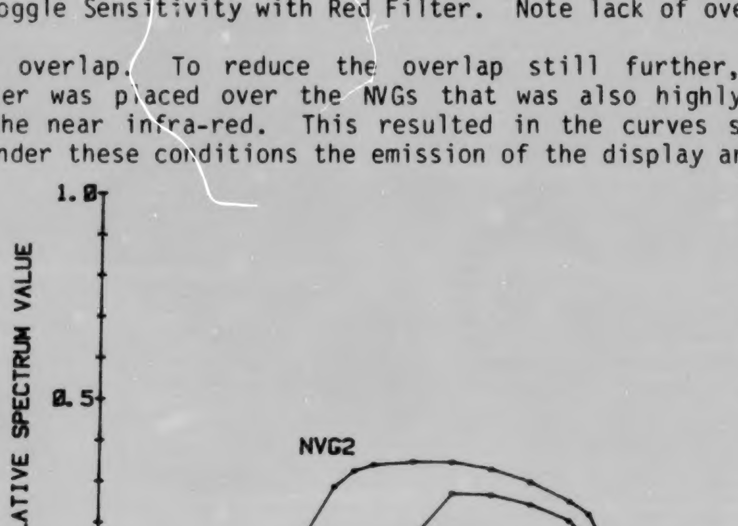


Figure 8. Comparison of Night Vision Goggle Sensitivity Without Filter (NVG 2) and With Red Filter (NVG 2 AA).

sensitivity of the NVG have almost no overlap ; thus effectively eliminating the interference of the display with proper operation of the NVGs. The red/infra-red filter over the NVGs also significantly reduced the sensitivity of the NVG to the blue-green E-L, thereby eliminating that source of interference.

The red/infra-red filter does reduce the overall sensitivity of the NVGs as shown in Figure 8. However, since most of the natural night illumination is in the near infra-red (IR), (i.e., 800 nm to 1,000 nm) the effective reduction in night sensitivity is very small.

Baffles for Light Control

It is not possible to turn off all incandescent lights in the cockpit since some are required instrument status lights. Several such lights were also unfortuantely located in the center console. Most of the center console was directly visible in the windscreen due to the reflection geometry. To control these reflections, a material developed by 3-M Corporation was applied wherever possible. This material, called Micro-Louver (ML), is like a miniature venetian blind cast in a thin plastic layer. It is about 1/16" thick and can be obtained in various configurations. By varying the "slat" spacing and tilt, the fan of light that is allowed through the material can be controlled. The material comes in three "fan widths" of 48°, 60°, and 90°; and several tilt angles: 0°, 18°, 30°, and 45°. The tilt angle refers to the direction with respect to the vertical. Thus a 48° fan at 0° tilt results in a light distribution that is emitted vertically (with respect to surface of material) with a 48° spread.

By placing appropriately chosen ML sections over the lights and displays, the light can be directed away from the windscreen toward the pilot, copilot, and flight engineer. This reduces or elimates the

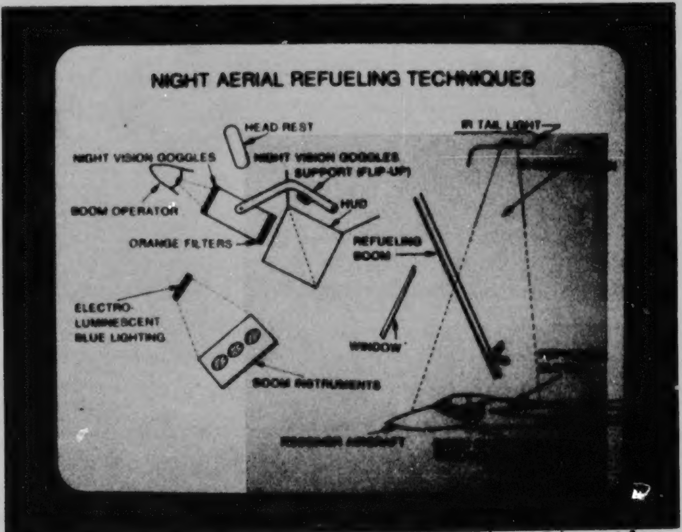

NIGHT AERIAL REFUELING TECHNIQUES
HEAD REST
IR TAIL LIGHT
NIGHT VISION GOGGLES
NIGHT VISION GOGGLES SUPPORT (FLIP-UP)
BOOM OPERATOR
HUD
ORANGE FILTERS
REFUELING BOOM
ELECTRO-LUMINESCENT BLUE LIGHTING
WINDOW
BOOM INSTRUMENTS
RECEIVER AIRCRAFT

Figure 9. The 48° Fan, 0° Tilt Micro-Louver View Directly.

direct reflections of instruments in the windscreen. Figure 9 shows a section of ML that is a 48° fan, 0° tilt mounted over a vu-graph. Note the clarity of the vu-graph beneath the ML. Figure 10 shows this same vu-graph and material but from a different angle. The vugraph behind the ML is almost completely black because the light has been directed in a fan upward.

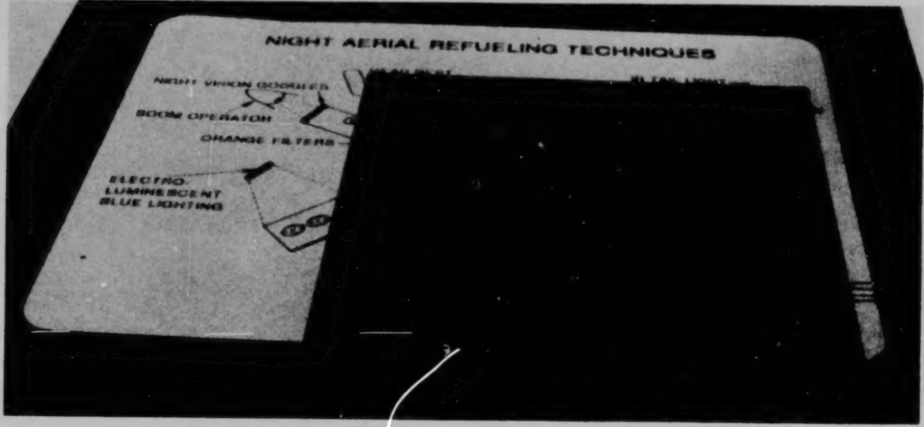


Figure 10. The 48° Fan, 0° Tilt Micro-Louver Viewed from an Angle of about 60°.

This technique of using ML baffles was successfully employed over several indicator lamps and displays including the incandescent lamp illuminated moving map display located in the center console. The ML was oriented to provide a horizontal fan of light directed away from the forward windscreen toward the pilot, copilot, and flight engineer positions.

Although this technique was highly successful for the visible light reflections, it was not totally successful with the IR reflections. The ML plastic was partially transmissive in the IR and the IR from the lamps was so strong that it caused scattering within the ML material. To combat this problem a thin, plastic material was borrowed from the laser safety industry. The original purpose of this material was to provide laser safety and protection at the near IR wavelengths. Thus the material passed a large portion of the visible spectrum but absorbed light in the near IR. This IR blocking material (IRBM) was used in conjunction with the ML material to provide fairly effective control of both visible and IR radiation. The IRBM was a "Glendale Green" filter material obtained from Glendale Optical Company. The published photopic transmissivity of the IRBM is about 45%. It does have a definite green tint and affects the red end of the visible much more severely than the green.

Anti-Flare Baffles on NVG

Another source of stray light that can affect the NVG operation is caused by flare. The NVGs have a 40° Field of View (FOV), 1:1 optical imaging system. However, bright light sources just outside of this FOV can still illuminate the objective lens of the NVG. Although this illumination is not imaged through the optical system (since it is outside the FOV) it still scatters in the lens causing a veiling luminance at the image plane that reduces contrast. This condition can be partly alleviated by providing an anti-flare baffle outside of the objective lens as shown in Figure 11. These baffles "shade" the objective lens from bright light sources outside of the FOV. They also provide a housing to mount the red/IR filters to the NVGs.

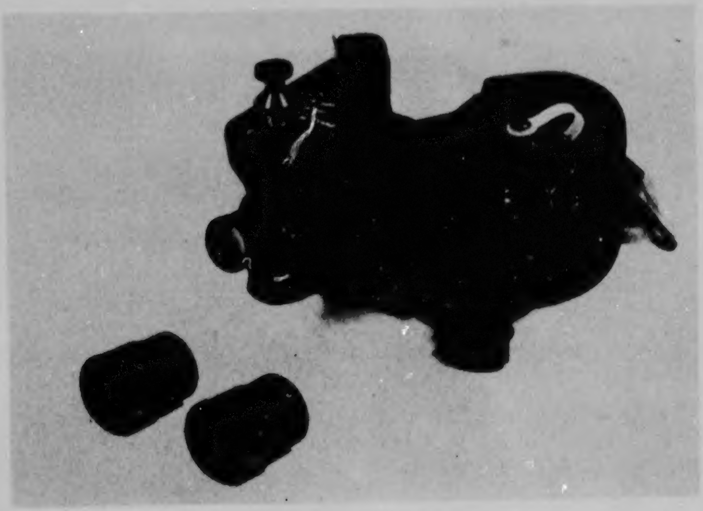


Figure 11. Anti-Flare Baffles and Filter Holders for the Night Vision Goggles.

Black Flight Clothing

To further reduce stray light it was recommended that the flight crew wear dark clothing to absorb any stray light instead of re-emitting it. Due to the geometry of the windscreen and pilot/copilot seats, the knees and hands of the pilot/copilot are reflected and highly visible in the windscreen to the NVG wearer. By wearing dark clothing the intensity of the these reflections were greatly reduced.

GROUND EVALUATION RESULTS

All of the light control techniques herein described were applied to the PAVE LOW III aircraft and evaluated by several instructor pilots during a day/night ground evaluation. Overall the evaluation results were extremely good. The copilot had sufficient light to do his job, but the lighting did not adversely affect the pilot's NVGs. Several specific problems were identified by the evaluating instructor pilots.

In general, the use of the IRBM tended to make some of the indicator lights too dim to the unaided eye in daytime. In particular, the moving map (navigation) display was marginally acceptable in daytime when sufficient IRBM was applied to block the IR emissions for night use. During the night evaluation several other sources of incandescent or neon light IR emissions were identified as requiring applications of the light control techniques. These sources were not identified originally because the other sources totally masked the light from these sources. However, with the original problem lights effectively controlled, these "secondary" sources of light control problems became evident.

Long term problems that need to be solved before these techniques can be applied are the materials problems associated with the ML and the IRBM. The ML is a soft plastic, as such it is susceptible to scratching and, in the case of the incandescent bulbs, it can be warped by the heat of the lamps. These same concerns apply to the IRBM as well.

If an aircraft cockpit used E-L panel lighting instead of the incandescent, then the heat problem associated with the ML application would be solved. Also, this would eliminate the need for the IRBM since the E-L emits no IR.

An additional bonus of an all E-L cockpit light system or application of the techniques described, is that the emission of IR from the cockpit is greatly reduced or eliminated. This should make the craft less visible and vulnerable to IR sensing and seeking devices.

CONCLUSIONS

The light control techiques described were successfully applied to the PAVE LOW III helicopter to make the interior lighting system compatible with the use of Night Vision Goggles. From the ground evaluation it is evident that these techniques provide a simple, inexpensive, and useful means to improve night visibility out of the cockpit with or without Night Vision Goggles. The materials problems encountered should be addressed, and these techniques should be considered in the design of new cockpit lay-outs for interior night lighting.

Since this effort, similar efforts have begun for light control of the C-130 and UH-1 aircrafts. These efforts are proceeding well as of this writing.

REFERENCES

Lees, Michael A.; Glick, David D.; Kimball, Kent A.; and Show, Allen C. Jr., In-Flight Performance with Night Vision Goggles During Reduced Illumination, U.S. Army Aeromedical Research Laboratory Report No. 76-27, Fort Rucker, Alabama, August 1976.

Sanders, Michael G.; Kimball, Kent A.; Frezell, Thomas L.; and Hofmann, Mark A., Aviator Performance Measurement During Low Altitude Rotary Wing Flight With the AN/PVS-5 Night Vision Goggles, U.S. Army Aeromedical Research Laboratory Report No. 76-10, Fort Rucker, Alabama, December 1975.

Task, Harry L. and Griffin, Leonard L., *Matched Filter Technique for Improving Visibility Through Windscreens Using Night Vision Goggles*, AF Invention #14,043, March 1980.

________________, *Electroluminescence: Lamps and Panels*, Grimes Division, Midland Ross Corporation report EL95 3/72, Urbana, Ohio.

PRELIMINARY EVALUATION OF AN ONBOARD COMPUTER-BASED INFORMATION SYSTEM

Sandra H. Rouse and William B. Rouse

University of Illinois
Urbana, Illinois 61801

A computer-based system for retrieval and display of procedural information will be discussed. Preliminary evaluation of this system will be considered in the context of two-person crews operating in normal, abnormal, and emergency situations. Data for execution times and errors will be reported for comparisons of computer-based displays and traditional hardcopy media. Further, the effects of various levels of computer "intelligence" will be reported.

INTELLIGIBILITY OF AND PILOTS' REACTIONS TO VARIOUS TYPES OF SYNTHESIZED SPEECH

G. Rothbauer
Technische Universitat Munchen

A literature search has shown that it would be of value to Voice Warning Research to clarify the question, whether an artificial or natural sounding computer-generated voice would be preferable for use in a cockpit voice warning system. Among the problems which arise when voice is used to communicate warnings are questions regarding the content and form of the information to be transmitted, interference with other acoustic information, priority logic and technical aspects of voice output.

Experiments will be conducted in a simple flight simulation using realistic cockpit noise and simulated radio communication. Pilots will be used as test persons: they are ideal because of their familiarity with general flight conditions as well as with phraseology used in aircraft communications. In order to guarantee an almost constant inter- and intraindividual operator workload, the difficulty of a two-dimensional control task - a simplified flying task - will be adaptively adjusted to individual performance capability.

The technical means used to generate voice warnings will be the independent variable in these experiments: one method will be electronic production of speech (not derived from spoken utterances) generated by a Votrax synthesizer; the other method will be digital storage of spoken warnings. A performance measure (intelligibility against a background of cockpit noise) as well as subjective reactions of the pilots will serve as dependent variables.

HEAD UP DISPLAYS IN OPERATION: SOME UNANSWERED QUESTIONS

Richard L. Newman

Crew Systems Consultants
Yellow Springs, Ohio

William L. Welde

Air Force Aerospace Medical Research Laboratory
Wright-Patterson AFB, Ohio

Because of the interest in using head up displays (HUDs) as primary flight references in instrument meteorological conditions (IMC), a survey of operational pilots flying HUD-equipped airplanes was undertaken. This survey revealed several problem areas. These are (1) The HUDs are too bright at night; (2) The field of view is too limited; (3) The location of the design eye reference point does not correspond to the typical pilot practice of sitting as high as possible to maximize his external view; (4) The response of the HUD symbols is not adequately controlled by existing specifications; (5) Pilots have an increased tendency towards spatial disorientation while flying by reference to the HUD; (6) The instrument landing system (ILS) displays are not satisfactory; and (7) The balance between presenting necessary information and clutter is not always achieved in today's HUDs.

Several recommendations are made to deal with these problem areas. Chief among these is the conducting of a flight experiment to develop design criteria for HUD symbol response and for the proper symbology for the ILS approach. These experiments would interact closely with the study of increased pilot disorientation when flying by reference to the HUD.

A STUDY OF DECISION-MAKING BEHAVIOR OF AIRCRAFT PILOTS DEVIATING FROM A PLANNED FLIGHT*

George W. Flathers, II
The MITRE Corporation
McLean, Virginia

and

Walter C. Giffin and Thomas H. Rockwell
Department of Industrial and Systems Engineering
The Ohio State University
Columbus, Ohio

ABSTRACT

This paper outlines an investigation into the worth structures of pilots facing a deviation from a planned flight. A "paper and pencil" simulation was used to frame the situation into which pilots interjected their own decision making skills in a simple ranking of candidate diversion airports with varying locational, navigational aid, radar and weather attributes. Using the conjoint measurement technique, attribute worth functions of thirty pilots were constructed. It was discovered that systematic differences in the worth functions of the pilots did not occur as a result of dividing the pilot sample according to any measure of flight hour experience. However, differences were found when the pilot sample was grouped according to grade of pilot certificate, type of pilot training, and type of flying most commonly done.

INTRODUCTION

Occasionally pilots must deviate from a planned flight and land at an airport which is not their ultimate destination or goal. The reason necessitating the change could be that the weather at the destination has turned bad, or that unexpected strong headwinds require another fuel stop. The reason could be more urgent in nature such as the failure of an aircraft system or subsystem which hinders some dimension of aircraft performance. In any case, the need to choose and to divert to another airport leaves the pilot with a decision, which, if not carefully made, could have serious consequences. This situation is called the diversion-decision problem.

---

*This paper is based on information developed as part of the Research Report "An Investigation Into Pilot and System Response To Critical In-Flight Events" by Thomas H. Rockwell and Walter C. Giffin, prepared for the NASA Ames Research Center under Contract NAS 2-10047.

In several accidents it can be pointed out that the most serious decisional deficiency occurred when the pilot was selecting a course of action in the wake of an unexpected event. In other words, once the pilot accepted the fact that a real or perceived problem situation existed, he made a poor decision concerning where to divert the flight. To understand why this happens, one should consider the sequence of steps a pilot must work through when resolving inflight problems.

Rockwell and Giffin (1980), in their study of pilot response to critical in-flight events, suggest that pilots respond to urgent situations in the four step process of detection, diagnosis, decision, and execution. Considering those situations in which a diversion is necessary, the decision step may be enlarged and dissected to reveal the pertinent influences in the diversion-decision process. One conceptualization of this breakdown is shown in Figure 1.

The pilot combines three main ingredients when making his diversion-decision. First, he uses his own diagnosis of the problem, including a reassessment of the needs, capabilities, and limitations of his airplane implied by the diagnosis. The quality of his diagnosis and reassessment can be measured on such dimensions as correctness and the amount of certainty he offers.

Secondly, the pilot brings in several psychological factors. Some of these are the economic, moral, social, and emotional motivating factors which oftentimes bias the overall decision process, even though they are rarely related to the objective evaluation of a situation. For example, the oft demonstrated "get home-itis" phenomenon may cause more rational alternatives to be ignored when faced with balky equipment and deteriorating weather.

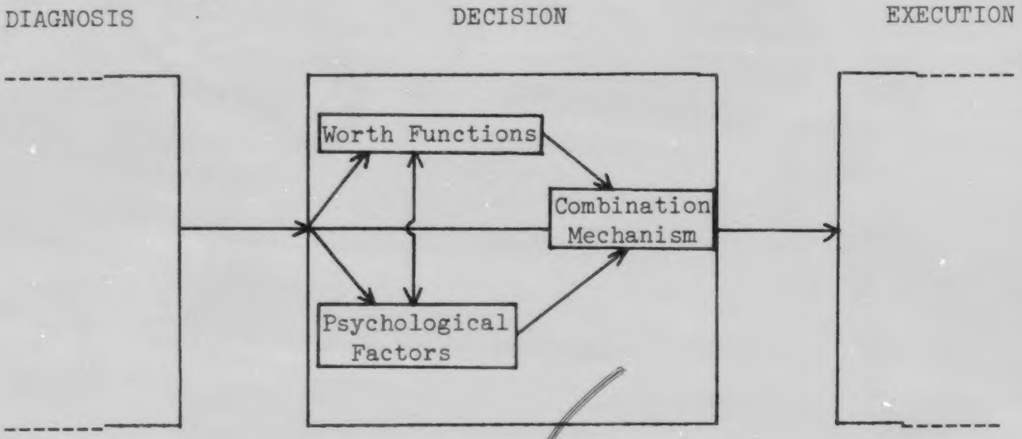


FIGURE 1. Internal Factors of Decision Step

Finally, the pilot incorporates a unique worth function, or preference function into the decision process. The worth function is the application of values or utilities to the decisional attribute. This function combines pieces of descriptive, physical information to yield the psychological "worth" of a decision alternative. Under homogeneous diagnostic and psychological conditions, the worth function determines how a decision will be made.

The focus of this study is on the worth functions pilots employ in the diversion-decision problem. Under the model of the decision process shown in Figure 1, the worth function could serve as a barometer of underlying psychological pressures in the decision process. It would be helpful to the understanding of the whole of pilot decision making processes if more was known about his worth function and its implications. Assessing the worth functions of a group of pilots could disclose differences in worth structures which are dependent on pilot background variables. For example, one might compare: experienced versus inexperienced pilots, conservative pilots versus risky pilots, and civil trained versus military trained pilots.

Only a small portion of available decision making literature has centered on decision making in the flight environment. Though pilot judgment has long been recognized as an important factor in the safety of flight, research in the area has largely been limited to observations made from outside the cockpit. Many of the observations are retrospective analyses of the decision making behavior of pilots involved in accidents or incidents. One reason more research has not been done can be traced to the difficulty in creating and controlling in-flight decision making environments.

A study by Jensen and Benel (1977) offers an excellent overview of pilot judgment from the viewpoints of aviation and psychology, and uncovers many of the aspects of this area which demand further research. Their main objectives were to examine popular questions about pilot judgment, offer a coherent definition of judgment, and determine whether training programs could be developed to modify judgment. A two-part definition of judgment was presented as well as evidence from judgment research in other fields that the process can be modified through training.

Hoffman, Slovic, and Rorer (1968) investigated the use of an analysis of variance model to determine how people combine various pieces of information in a decision making situation. Nine radiologists were asked to rate the degree of malignancy of ninety-six hypothetical ulcers on a scale of one to seven. The ulcers were described in terms of six orthogonal dimensions ("orthogonal" meaning changes in value along one dimension do not imply changes on other dimensions). An ANOVA was also used to determine the significance of configural use of attribute information (i.e., combining the attributes of an alternative interactively). They found distinct difference in the decision models used by the different radiologists, but also found that, on the average, about 90% of a subject's response variation could be attributed to main effects only.

In those cases where measurement of some of the values of the worth model is difficult, the conjoint measurement approach has been useful. It provides a stepping stone between those decision tasks with hard-to-measure variables and the ANOVA method of processing data which is more easily quantified. By doing so, much can be inferred about the process governing a decision situation which would otherwise not be approachable. As a product of the field of mathematical psychology, conjoint measurement is the topic of numerous articles in the behavioral sciences. A thorough review of conjoint measurement can also be found in Green and Wind's (1973) book.

Conjoint measurement has been used in an aviation decision making study by Curry (1976). He determined the worth functions for general aviation and airline pilots who were considering several landing situations. One part of his study revealed large differences between the worth functions of a general aviation pilot and an airline pilot.

ANALYZING THE DECISION MAKING PROCESS

The pilot in the diversion-decision problem is faced with trade-offs in matching available airports against his concept of "ideal". The "ideal" airport he has in mind may not be a feasible alternative for his situation. The pilot faces what is known as a multiattribute (or multidimensional) decision problem. To understand how these decisions are made requires ascertaining how one trades off conflicting criteria, for example, better weather versus an increase in distance or time. That is the central issue of this research effort.

Many researchers have been interested in determining the worth functions people use in choice making situations. There are problems, though, that prevent the use of a straightforward analysis, such as regression, in determining a worth function. These problems arise as a result of the difficulty in measuring the value of some of the variables. However, by carefully arranging the alternatives in the choice making situation, one can overcome these problems by using the conjoint measurement approach. This is possible because conjoint measurement requires only the subjective rank of a set of alternatives instead of absolute worth values.

Although a relatively recent concept, conjoint measurement has found widespread application in many different fields. Examples of its application can be found in marketing research (Johnson, 1974), psychology (Krantz and Tversky, 1971), and mathematical psychology (Tversky, 1967).

Conjoint measurement has been used in aviation decision making research in assessing the worth of an approach to landing (Curry, 1976). Curry used conjoint measurement to determine the relative worth of attributes in the overall preference function of general aviation pilots and airline pilots who were considering an approach-to-landing situation. The variables he considered were wind direction, wind velocity, runway surface conditions, and turbulence.

The main concern of conjoint measurement is to determine the joint effect of two or more independent variables on the ordering of a dependent variable. From the viewpoint of multiattribute decision making, the additive model (where the total worth is simply a sum of the partial worths of each attribute) appears the most relevant in most situations. The goal of conjoint measurement, is to find a set of partial worths (interval scaled with common unit), one for each attribute in the worth function, such that the ordering of the sums of the partial worths preserves, as closely as possible, the ordering of the dependent variable.

When considering the additive model for worth functions, an analysis of ordinally rank alternatives (on the basis of preference, for example) can proceed by using a regression approach of the analysis of variance. Assuming monotonic partial worth functions, a main-effects analysis of variance is performed. The independent variables take on the values which are used to define the attributes or dimensions of the alternatives. If, for example, attribute i was allowed to vary on three levels, the corresponding values of the independent variable could range from, low to high, -1 to 0 to +1. The coefficients of the additive worth model, which reveal how much emphasis the attributes have in the decision making situation, are obtained by performing a regression analysis over the whole set of ranked objects. The values of the dependent variables are taken from the rank position of the various choices. This substitution says, in effect, that the worth values (dependent variable) for all the alternatives are at equal intervals along the worth scale. Though this assumption is very likely untrue, it provides an ideal place for the regression to begin. In fact, several researchers have found that rarely do other values for the dependent variables yield the coefficients in regression which make the worth function fit the observed data better than when rank values are used (Curry, (1976), and Green and Wind (1973)).

This analysis yields a function which very closely approximates a subject's preference scheme by considering simple, ordinally ranked data. The function can then be held up for observation to disclose which attributes are dominant in the subject's preference scheme or worth function. If the worth functions of several subjects are calculated under homogeneous conditions, an analysis can be performed to determine how properties of preference functions might be related to background variables of the subjects.

THE EXPERIMENT

Thirty instrument rated pilots volunteered to participate in a joint workshop-experiment on pilot response to critical in-flight events and decision making. Prior to any of the individual experimental sessions, the participants filled out a pilot background data form on which their flight experience, type of flight training, and type of flying most commonly done was recorded. The pilots also recorded their responses to several items on a knowledge survey which was designed to estimate the pilot's knowledge of general aircraft systems. The pilot-subjects were briefed on general weather and aircraft information which would be useful in the "paper and pencil" simulations which were to follow.

On a one-to-one experimenter-subject basis, the pilots were asked to consider five different scenarios unveiled through paper and pencil techniques, one at a time. The first four scenarios were designed to let the pilot exercise his diagnostic skills on airplanes which were experiencing problems in flight. The last scenario was designed to reveal the pilot's decision making tendencies in the diversion-decision problem. It was the last scenario that structured the platform on which the data for this study was gathered. The paper and pencil simulation began when the experimenter commenced reading a chronologically based description of the flight beginning with the preflight planning stages. The opening paragraph described the mission of the flight, and provided some general preliminary information:

> You are at the Bangor International Airport in Bangor, Maine, and desire to fly to Glens Falls, New York, for a 1:00 p.m. business meeting (refer to the Low Altitude Enroute chart provided). The current time is 9:00 a.m. and you feel you can be ready for departure by 10:00 a.m. after you conduct all necessary preflight activities. The plane you will be flying today is your company's Cherokee Arrow (N8086W). You have flown this particular plane several times before and regard it as a reliable airplane. The aircraft's fuel tanks are full, and after a very thorough preflight inspection you conclude that it is operationally and legally ready for the flight. A brief list of the important performance figures and IFR equipment on board is shown in the table provided.

The subject was given a moment to review the appropriate tables and figures. Because there was no element of actual control (e.g., control of aircraft attitude, control of engine output, etc.), in the paper and pencil simulation, the information in these documents provided the basis for complete planning of the hypothetical flight. From this information, and the information he would receive on the weather, the pilot was able to prepare for all phases of the flight.

In the interest of saving time and reducing variations in pilot estimates of such flight performance variables as time enroute, an IFR flight plan was computed and filled out for him. The pilot was given a few moments to check over the flight plan and accept it.

In the actual flight portion, the text described to the pilot the takeoff and departure of his flight, the climb to cruising altitude, and the initial phases of cruise flight. The pilot was encouraged to follow the progress of his flight on a simplified version of the Low Altitude Enroute chart for his route of flight. Until the midway point in the flight, every aspect of the operation was considered routine and according to schedule.

About midway through the flight, the need to divert was introduced. In this case it took the form of an alternator failure. The symptoms of the problem were stated, some corrective actions were taken, and a general assessment of the situation was provided in the text which follows:

> At 11:21 (one hour and twenty-one minutes after departure) you cross Grump intersection. One minute later you hear a short burst of static noise over your radio speakers. At the same time you notice your VOR needles and their "on-off" flags flicker unsteadily and return to normal indications. Curious to know what caused these events, you glance over the instrument panel and find a "zero" reading on the ammeter. You actuate the landing light and notice no change in ammeter indications. From this information, you conclude the alternator has failed. You follow the procedures in the manual but your attempts to bring the alternator back into service are unsuccessful. Therefore, you turn off the alternator, minimize the electrical load, and operate solely on battery power.
>
> The battery, by itself, can supply the required power to operate your radios for only a limited time. The amount of time you have depends on the size and condition of the battery, the temperature, and the power requirements of the essential electrical equipment you use. Even under ideal conditions battery power is not expected to last longer than fifty minutes.
>
> You are at an altitude of 8000 feet, just west of Grump intersection. The time is now 11:23 and you have been airborne for one hour and twenty-three minutes. Winds are out of the southwest at thirty knots.

The preceding text contains several key pieces of information for the pilot facing the diversion-decision problem. First, the symptoms and the diagnosis set the stage for the need to divert. The straightforward statement of the diagnosis was intended to give each pilot the same perception of the problem. This is of great importance since the focus of this study is on the diversion-decision issue rather than diagnosis. If left to their own diagnostic devices, it would be highly unlikely that all pilots would perceive the problem the same way. Next, the ramifications of the problem were clearly assessed. Having only battery power left to run electrical equipment the problem was made urgent in terms of time. The consequences of flying beyond the lifetime of the battery are serious: the flight would be trapped aloft with no means of communication or navigational guidance. Finally, an estimated maximum time the battery would be useable was stated as being not longer than fifty minutes.

The paper and pencil simulation had now reached the point where the pilot was called upon to inject his own personal skills into the

142

diversion-decision problem. It involved ranking a group of airports from most preferable to least preferable based on their attributes.

There were four attributes considered at each of the sixteen airports in this ranking task, namely: (1) air traffic control (ATC) services at the airport, (2) the weather at the airport, (3) the time to fly from present position to the airport, and (4) the best instrument approach facilities there. Because the number of alternatives to consider has to increase dramatically as the number of attributes increases, it was desired to keep the number of attributes small. The selected attributes were chosen because they are orthogonal with respect to one another, and they were the more pertinent items to consider in this situation.

Each attribute was varied over two levels. Four attributes at two levels required sixteen candidate airports each with a different set of attributes. See Table 1 for the experimental design and a description of attribute levels.

Each alternative airport was depicted on a three by five inch card in terms of the four attributes. The sixteen cards were shuffled (prior to the experiment) and laid out before the subject in a random fashion. The subject was then asked to rank the airports from most preferable to least preferable, given the situation he was in.

Subjects were given as much time as they needed to complete the ranking task, but rarely did it take longer than five minutes. Subjects generally appeared quite involved and made meticulous adjustments to the rank before yielding a final ordering. When the subject had completed the ranking task and was satisfied with the final product, the experimenter recorded the sequence.

ANALYSIS AND RESULTS

The analysis proceeded with the ranks the subjects provided in the experimental task. The first part of the analysis was aimed at modeling the pilot's worth function based on the ranked data. Conjoint measurement was used in this part. The second part of the analysis combined the results of the first part and background data on the pilot to determine if worth functions are related to background variables and diagnostic performance.

The convention adopted for this study was that a rank value of sixteen was the most preferable airport and a rank value of one was least preferable. Most of the pilots agreed airport A of Table 1 was most preferable and airport P was least preferable. However, much variation was seen in the airports in between.

The first step was to determine the coefficients for each subject's worth function. The general form of the worth function under consideration is of the simple additive type described in Equation 1.

TABLE 1

2^4 Factorial Layout of Airports Attributes

Airports	Attributes			
	ATC	Weather	Time	Approach
A	+	+	+	+
B	-	+	+	+
C	+	-	+	+
D	-	-	+	+
E	+	+	-	+
F	-	+	-	+
G	+	-	-	+
H	-	-	-	+
I	+	+	+	-
J	-	+	+	-
K	+	-	+	-
L	-	-	+	-
M	+	+	-	-
N	-	+	-	-
O	+	-	-	-
P	-	-	-	-

(+ = High Value; - = Low Value)

Key: ATC: + = Tower with radar
- = UNICOM

Time: + = 15 minutes to reach the airport
- = 30 minutes to reach the airport

Weather: + = 1000' ceiling, visibility 3 miles
- = 500' ceiling, visibility 1 mile

Approach: + = Instrument Landing System (ILS)
- = Non-Directional Beacon (NDB)

$$W(Z) = BATC*XATC + BWX*XWX + BTIM*XTIM + BAPP*XAPP, \tag{1}$$

where XATC, XWX, XTIM, and XAPP are the independent variables describing airport Z in terms of air traffic control services, weather, time, and approach aids, respectively, and BATC, BWX, BTIM, and BAPP are the respective coefficients. In this study the independent variables took on the value of +1 or -1, depending on the level of that variable (attribute) at airport Z. In order to begin a regression analysis, the actual rank position for airport Z was used in place of the (still unknown) worth value for Z.

After the data was properly organized and coded, it was submitted to a computer program for regression analysis. The purpose of the regression was to come up with estimates for the four B coefficients of the worth function.

The range of values for the coefficients was 0.250 to 4.000. An interpretation of the coefficients can be given if one considers the entire worth function for each subject. The coefficients for one subject, for example, were 1.000, 2.000, 4.000, and 0.500 for ATC, weather, time, and approach, respectively. This can be interpreted as follows: the worth of having the time value at the high level (fifteen minutes) was twice that of having the weather value at the high level (1000 feet ceiling, three miles visibility), four times the worth of having ATC at the high level (tower with radar), and eight times the worth of having the approach at the high level (ILS). In other words, the most important feature about each airport for this subject was time followed by weather, ATC, and approach, in that order.

A check of the goodness-of-fit of the derived worth function to the input rank was performed. The Spearman rank correlation coefficient was calculated for the relationship between the computed worth values for the airports and the subject's input rank. The lowest Spearman coefficient observed was .874, and all but two were above .930. The input ranks of five subjects were perfectly correlated to the computed worth values resulting in a Spearman coefficient equal to 1.000.

In addition to the correlation analysis, an analysis of variance was performed to reveal what portion of the subject's response variation could be attributed to the model. It was found that, for all subjects, well over 90% of the response variation was accounted for by the model itself. When interaction terms were allowed to exist (as in a linear-interactive model), their resulting contributions to variation were found to be on the order of five to thirty times smaller than the additive terms.

All of these findings offer convincing evidence that the additive model was, indeed, the appropriate choice for this decision making situation. Having this confidence in the derived worth function, it was then possible to analyze differences in the worth functions and how they relate to pilot background variables.

The basic approach was to dichotomize the sample population based on several different descriptors of a pilot's background and skill. The splits were performed on the basis of flight experience, training, type of pilot certificates, type of flying most commonly engaged in, and measures of ability determined by the knowledge survey and other means. The worth coefficients from the regression analysis became the center of attention in this analysis. Table 2 summarizes the results of various groupings and highlights those which are statistically significant.

TABLE 2

Summary of Coefficient Means By Pilot Category*

No. of Subjects	Pilot Category	Group Coefficient Means							
		$\overline{\text{BATC}}$		$\overline{\text{BWX}}$		$\overline{\text{BTIM}}$		$\overline{\text{BAPP}}$	
		I	II	I	II	I	II	I	II
 (22) (8)	Total Flight Hours: Category I: Time > 1100 hrs. Category II: Time < 1100 hrs.	1.74	1.90	2.08	2.54	1.80	1.90	1.91	1.41
 (22) (8)	Total Single Engine Hours: Category I: Time > 800 hrs. Category II: Time < 800 hrs.	1.91	1.44	2.05	2.66	1.74	2.04	1.91	1.41
 (14) (16)	IFR Hours: Category I: Time > 300 hrs. Category II: Time < 300 hrs.	1.57	1.94	2.13	2.28	2.00	1.68	1.91	1.70
 (7) (23)	Type of Training: Category I: Military Training Category II: Civil Training	1.25	1.92	2.40	2.16	2.20	1.72	1.68	1.90
 (8) (22)	Grade of Certificate: Category I: Airline Transport Category II: Private or Commercial	1.73	1.80	1.48	2.42	2.53	1.56	1.73	1.80
 (5) (25)	Type of Flying Mostly Done: Category I: Military or Airline Category II: Pleasure, GA/Comm, Business	0.90	1.96	2.20	2.21	2.55	1.67	1.74	1.90

*Significant Differences Enclosed In Dashed Lines

The first dichotomization was performed on the basis of total flight experience in terms of flight hours. The sample was split at the natural break nearest the 50th percentile. The criterion used to split the pilots was 1100 hours. Nine pilots were in the lower category and twenty-one pilots in the higher category. A t-test was performed to determine if there were any significant differences between the means of the B coefficients for the two groups. At the p = .10 level, no significant differences were found. Changing the criterion in either direction had no effect. Splits of the sample were also made based on the number of Instrument Flight Rules (IFR) hours and single engine airplane hours. At the .10 level, no significant differences were found.

The type of training a pilot received was used as a criterion to split the sample. There were seven military trained pilots and twenty-three civil trained pilots. A t-test was performed on the worth coefficients and a difference which was significant at the .10 level (p = .06) was observed for the mean value of BATC. (Recall that BATC is a measure of the importance of air traffic control facilities in airport worth evaluation.) For civil trained pilots the mean value of BATC was 1.92 and for military trained pilots it was 1.25.

The pilot sample was split on the basis of the type of certificate the pilot held. In this case, the twenty-one pilots with Private and Commercial certificates made up one group, and the nine pilots with Airline Transport Pilot certificates made up the other group. Some notable differences were observed when comparing the worth coefficients of these two groups. The mean value of BWX, a measure of the importance of weather to a pilot in this situation, was 2.49 for private and commercial pilots and 1.48 for airline transport pilots. A t-test was performed and this difference was found to be significant (p = .05). Another difference, significant at the .10 level, was observed for the value of BTIM, a measure of the importance of the time attribute. For airline transport pilots BTIM had a mean value of 2.53 and for private and commercial pilots the mean value was 1.56.

The type of flying most commonly done was also used as a basis to divide the pilot sample. Pilots who engaged primarily in business, light commercial, or pleasure flying made up one group. Pilots who were involved with airline or military flying comprised the other group. The split was made in this fashion because the highly structured environments in which airline and military pilots operate are similar in many ways. They are both usually required to fly in and out of busy terminals and heed schedules, policies, and other disciplines. Pilots who fly for business, light commercial, or pleasure concerns, however, operate in a much more relaxed atmosphere and dictate their own policies. Based on this split of the sample population, a significant difference (p = .024) was observed for the coefficient BATC. The mean value for business, light commercial, and pleasure flyers was 1.96, while the value for airline and military pilots was 0.90.

CONCLUSIONS

In summary, significant differences in worth function coefficients were not a result of flight hour experience in any category. The most significant differences were related to the grade of pilot certificate, the amount and type of initial and recurrent training, and the type of flying most commonly done. This suggests that training and repeated exposure to testing situations are the variables which can predict the general form of a pilot's worth function. A closer examination of the training and certification process is in order.

Pilots with higher levels of skill (as implied by higher grades of certificates, more regimented training, and repeated exposure to demanding flying situations) placed less emphasis on ATC, weather, and approach aids. This was interpreted as an indication that the lower values of these attributes were still within the realm of these pilots' self-perceived skill level. As a result, more emphasis could be placed on time. Pilots who were not in the higher skill category, on the other hand, could not consider some of the airports with the same amount of confidence. Consequently, time sacrifices had to be made.

Once a pilot obtains a private pilot certificate with an instrument rating, there is little he must legally do to continue exercising the privileges of his certificate. He can continue to accrue many hours of flight time but he is required to demonstrate, on only a sporadic basis, that he is maintaining his basic skills. Airline transport pilots and those pilots who fly for the military or airlines, however, must maintain a higher level of skill regardless of the amount of flight time they have. The general level of preparedness is much higher for military and airline flyers than for the rest of the flying population. All of this lends support to the notion that the total amount of flight experience is not as important as the amount and quality of initial and recurrent training in determining the general worth structure of a pilot.

An important issue supported by this research is that pilots can infer worth functions. There are differences in the way the thirty pilot subjects view worth functions. Although differences were found it is sometimes difficult to explain the logic of why they are different.

REFERENCES

1. Curry, R., "Worth Assessments of Approach to Landing", *Proceedings of the Annual Conference on Manual Control*, 1976, 585-591.

2. Green, P., and Wind, Y., *Multiattribute Decisions in Marketing*, Hinsdale, Illinois: Dryden Press, 1973.

3. Hoffman, P.J., Slovic, P., and Rorer, L. G., "An Analysis of Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment", *Psychological Bulletin* 1968, *5*, 338-349.

4. Jensen, R. S., and Benel, R. A., *Judgment Evaluation and Instruction in Civil Pilot Training*, Springfield, Virginia: National Technical Information Service, FAA-RD-78-24, 1977.

5. Johnson, R. M., "Trade-Off Analysis of Consumer Values", *Journal of Marketing Research*, 1974, 11, 121-127.

6. Krantz, D. H., and Tversky, A., "Conjoint Measurement Analysis of Composition Rules in Psychology", *Psychological Review*, 1971, 2, 151-169.

7. Rockwell, T. H., and Giffin, W. C., "An Investigation of Pilot and System Response to Critical In-Flight Events", on-going research at The Ohio State University under NASA contract, 1980.

8. Tversky, A., "A General Theory of Polynomial Conjoint Measurement", *Journal of Mathematical Psychology*, 1967, *4*, 1-20.

AN ANALYSIS OF AIRCREW PROCEDURAL COMPLIANCE*

Jeffrey E. Schofield
Department of Mathematical Sciences
U.S. Air Force Academy
Colorado Springs, Colorado

and

Walter C. Giffin
Department of Industrial and Systems Engineering
The Ohio State University
Columbus, Ohio

ABSTRACT

This research examines the relationships between aircrew compliance with procedures and operator errors. The data for this analysis were generated by reexamination of a 1976 experiment in full mission simulation conducted by Dr. H. P. Ruffell Smith for the NASA-Ames Research Center. The character of individual operators, the chemistry of crew composition, and complex aspects of the operational environment affected procedural compliance by crew members. Associations between enumerated operator errors and several objective indicators of crew coordination were investigated. The correspondence among high operator error counts and infrequent compliance with specific crew coordination requirements was most notable when copilots were accountable for control of flight parameters.

INTRODUCTION

An analysis of aircrew procedural compliance was undertaken as part of a larger project at The Ohio State University titled "An Investigation Into Pilot and System Response To Critical In-Flight Events" (1). The project was sponsored by the NASA Ames Research Center. The purpose of this portion of the project was to examine airline cockpit crew operations in terms of the influence of procedural compliance on observed operator errors. A more complete discussion of this material can be obtained from Schofield's Ph.D. Dissertation (3).

The study was based upon data generated in an experiment conducted in 1976 by Dr. H. P. Ruffell Smith under the auspices of the NASA-Ames Research Center (2). The Ruffell Smith research utilized a full-mission simulation to study the performance of fully qualified airline crews under varying conditions of workload. The cockpit was that of a Boeing

---

*This paper is based on information developed as part of the Research Report "An Investigation Into Pilot and System Response To Critical In-Flight Events" by Thomas H. Rockwell and Walter C. Giffin, prepared for the NASA Ames Research Center under Contract NAS 2-10047.

747 which accomodated the usual three-person crew plus two observers, a simulator operator/traffic controller, and an audio coordinator. The full-mission scenario used was built around a charter flight from Dulles Airport to Heathrow Airport (London) with a thirty-minute intermediate stop at Kennedy Airport (New York) for fuel and cargo. The first segment placed relatively low workload on the crews, while the second segment was much higher due to pre-programmed mechanical failures.

Ruffell Smith concentrated on crew errors during the second segment (high workload) of the scenario. He was interested in establishing statistically significant physiological or historical predictors of crew performance during the second leg. This study, on the other hand, emphasized the routine or customary tasks of flight operations as exemplified by the first segment of the Ruffell Smith scenario. Furthermore, it was concerned with: 1) quantifying routine crew procedures, 2) analyzing observed crew errors to identify which particular crew members were the primary causes of such errors, and 3) comparing measures of procedural compliance and operator error.

The primary data came from the audio tracks of the FM tapes and handwritten documents generated by the Ruffell Smith study. This information was supplemented by data which was culled from the Aircraft Operating Manual, the Company Operations Manual, the Federal Aviation Regulations, crew handbooks, and assorted navigational documents.

RESULTS

The principal results of this study include enumeration of normal operating procedures, an assessment of aircrew compliance with certain of those procedures and an analysis of operator errors related to procedural compliance.

Procedures

A procedure is defined as "a symbolic and mnemonic representation of a set of sensory, cognitive, and/or motor activities which, when recalled and executed within determinable tolerances, complete a task as designed". The word "procedure" and its many aliases appear throughout aviation literature. Nineteen separate words and phrases associated with aircrew operations which have procedural connotation are identified.

A set of normal operating procedures, as opposed to Abnormal, Alternate, Irregular, or Emergency procedures, are enumerated which represent an idealized sequence based on the events in the Dulles-JFK segment of the Ruffell Smith experimental scenario. All of these procedures are considered mandatory for normal flight operations in instrument meteorological conditions. Each procedure is identified by published format and the cockpit crew members expected to exhibit active procedural behavior. These are catalogued in Table 1.

The astonishing fact in this list is that 97 normal operating procedures can be identified for standard cockpit activities lasting

15?

TABLE 1

Normal Operating Procedures

Index Number	Name	Format(1) codes	Operator(2) codes	Crew(3) coordination
1.	Basic ATC communications practices	N	PNF	
2.	Preflight Radio Checklist	C,N	P2	
3.	Gear pin status report	N	P1	
4.	Hydraulic system pressurization	N	P1 & FE	
5.	ATIS report	N	U	
6.	Clearance Delivery communications	N	PNF	
7.	Ground Control communications	N	PNF	
8.	Pre-start Checklist	C,N	A	x
9.	Ground crew report	N	P1	
10.	Cabin report	N	P1	
11.	Engineer's Start Checklist	C,N	FE	
12.	Start Checklist	C,N	A	x
13.	Engine starting	N	P1 & FE	
14.	Ground connections and hand signals report	N	P1	
15.	Engineer's Taxi Checklist	C,N	FE	
16.	Pre-taxi Checklist	C,N	A	x
17.	Transfer of EGT monitor	N	P1 & FE	x
18.	Ground Control communications	N	PNF	
19.	Taxi	N	PF	
20.	Takeoff and departure briefing	N	PF	
21.	Final weight and balance computation	G,N	FE	
22.	Taxi Checklist	C,N	A	x
23.	Tower communications	N	PNF	
24.	Passenger pre-takeoff announcement	N	PF	
25.	Engineer's Takeoff Checklist	C,G,N	FE	
26.	Runway line up	N	PF	
27.	Takeoff Checklist	C,N	A	x
28.	Thrust setting (takeoff power)	G,N	PF & FE	
29.	Takeoff	N	PF	
30.	Takeoff callouts	N	PNF	x
31.	Noise abatement departure	G,N	PF	
32.	Gear retraction	N	PF & PNF	x
33.	Departure Control communications (initial contact)	N	PNF	
34.	Thrust setting (rated power)	G,N	PF & FE	
35.	Departure Control communication (radar vector)	N	PNF	
36.	Flap retraction	G,N	PF & PNF	x
37.	Altitude callout	N	PNF	x
38.	Intermediate level off	N	PF	
39.	Departure Control communications (climb clearance)	N	PNF	

TABLE 1 (con't.)

Index Number	Name	Format codes (1)	Operator codes (2)	Crew coordination (3)
40.	Airways navigation practices	N	PF & PNF	
41.	Thrust setting (rated power)	G,N	PF & FE	
42.	Climb (below 10,000 feet MSL)	N	PF	
43.	Company departure report	N	U	
44.	ARTCC communications (initial contact)	N	PNF	
45.	Seat belt sign	N	U	
46.	Climb (above 10,000 feet MSL)	N	PF	
47.	After Takeoff Checklist	C,N	PF & FE	
48.	Altimeter reset (not applicable for cruising below 18,000 feet)	N	A	
49.	ARTCC communications (route clearance)	N	PNF	
50.	Cruise data	G,N	FE	
51.	Altitude callout (climb)	N	PNF	x
52.	Level off	N	PF	
53.	Mach number/airspeed crosscheck	N	FE	
54.	Cruise	N	PF	
55.	ARTCC communications (radar vector)	N	PNF	
56.	Turbulence penetration	N	A	
57.	ARTCC communications (radar vector)	N	PNF	
58.	Turbulence exit	N	A	
59.	ARTCC communications (route clearance)	N	PNF	
60.	Fuel systems management	G,N	FE	
61.	ARTCC communications (center change; initial contact)	N	PNF	
62.	ATIS report	N	U	
63.	Company arrival report	N	U	
64.	Approach briefing	N	PF	
65.	ARTCC communications (sector change; initial contact)	N	PNF	
66.	Approach data and speed bugs	G,N	A	
67.	Passenger arrival announcement	N	PF	
68.	Descent Checklist	C,N	A	x
69.	ARTCC communications (descent clearance)	N	PNF	
70.	Descent (above 10,000 feet MSL)	N	PF	
71.	Altimeter reset	N	A	
72.	Seat belt sign and landing lights	N	U	
73.	Descent (below 10,000 feet MSL)	N	PF	
74.	Approach Control communications (initial contact; clearance)	N	PNF	
75.	Approach Checklist	C,N	A	x
76.	Category I Instrument Landing System (ILS) Approach	G,N	PF	

TABLE 1 (con't.)

Index Number	Name	Format(1) codes	Operator(2) codes	Crew coordination (3)
77.	Approach radio checks	N	PF & PNF	
78.	Altitude callout (descent)	N	PNF	x
79.	No smoking sign	N	U	
80.	Approach Control communications (radar vector)	N	PNF	
81.	Approach flap extension	G,N	PF & PNF	x
82.	Course bar and glide slope callouts	N	PNF	
83.	Approach Control communications (approach clearance)	N	PNF	
84.	Landing gear/landing flap extension	G,N	PF & PNF	x
85.	Landing Checklist	C,N	A	x
86.	Final approach fix (FAF) Communications	N	PNF	
87.	FAF instrument crosscheck	N	PNF	
88.	Precision approach callout	N	PNF	x
89.	Outside scan and visibility callouts	N	PNF	
90.	Landing	N	PF	
91.	Landing roll callouts	N	PNF & FE	x
92.	Tower communications	N	PNF	
93.	After Landing Checklist	C,N	A	
94.	Taxi	N	PF	
95.	Ground Control communications	N	PNF	
96.	Parking	N	PF	
97.	Blocks Checklist	C,N	A	
98.	Transfer of Aircraft Control(4)	N	PF & PNF	x

(1) Format Codes: C (Checklist), G (Graphical), N (Narrative)

(2) Operator Codes: A (All), P1 (Captain), P2 (Copilot), FE (Flight Engineer), PF (Pilot Flying), PNF (Pilot Not Flying), U (Unspecified)

(3) Procedures used for quantitative compliance assessments

(4) This is an optional procedure which is to be used at the discretion of the pilot flying in the aircraft

approximately 75 minutes. This lengthy list does not include any "optional" procedures or emergency type procedures. They represent only standard operating procedures for the first leg of the simulated flight scenario.

Several empirical taxonomies have been developed which seek to classify these procedures in ways to identify useful relationships among them. One such grouping is the set of 21 crew coordination procedures shown in Table 1. Crew coordination procedures are emphasized since they capture the essential ingredients of group leadership, crew management and behavioral conformity. This study examines the relationships between meticulous compliance with coordination procedures and the crew errors noted by Ruffell Smith.

Compliance Assessment

Although Ruffell Smith used eighteen crews in his experiment, the quality of the data generated and the observers were not the same for all eighteen simulation runs. Ten runs were selected, which had the same set of observers and usable audio data throughout, for detailed procedural analysis. The 21 crew coordination procedures were further subdivided into checklists, callouts, configuration changes, and transfers. Performance of each of the ten crews was then evaluated for each subdivision.

Pre-start, Start, Pre-Taxi, and Takeoff Checklists are supposed to be initiated upon command of the captain or the flying pilot. The other pilot is then to announce the name of the checklist as a confirmation of the command, and read the opening challenge. Once initiated, checklists may be delayed by interruptions, but ultimately must be resumed and completed in toto.

In every experimental run the requisite challenges and responses were made, even though some of the operator actions and replies were contrary to procedural specifications. However, there were remarkable differences in the patterns of behavior noted among crews for these five checklists. In a total of fifty opportunities over ten flights, the command-announcement-challenge sequence was fully executed only five times. The observed shortcuts raised questions about possible degradation in crew cohesion leading to increased uncertainty and lack of internal order.

The five audible checklists conducted by the two pilot crewmembers, were contrasted with three checklist sequences (Descent, Approach, and Landing) in which the flight engineer was the challenger. Exactly half of the observed thirty sequences here began in the prescribed command-announcement-challenge order and only one was missing the initial command. In addition to collectively making more of the prescribed announcements than their pilot counterparts, the flight engineers were more self-consistent. Three engineers omitted all announcements and three others omitted one. They also were more consistent than pilots in following the response to the last challenge statement with the prescribed procedure completion statement. When a pilot was the last challenger, 20% of the time the completion statement was omitted; when an engineer was the last challenger, only 4% were omitted. One plausible hypothesis is that crew coordination might be improved by making the flight engineer the challenger of all checklists.

Callout procedures are fundamentally different from checklists. In the usual format the non-flying pilot acts as a back-up or second-level visual monitor who audibly relays operating information to the flying pilot. Callouts occur during take-off, climb, descent, approach, and landing.

One hundred seventy opportunities, among the ten crews, to execute callout procedures were identified. Thirty-eight procedural errors were noted, half of which were errors in altitude callouts during climb or descent. The errors noted were callouts made by the flying pilot rather than the non-flying pilot (seven cases), late callouts (thirteen cases), and omitted callouts (twenty cases).

Procedures for gear and flap extension/retraction were well executed in terms of established oral procedures. In 104 observed configuration changes one of the two prescribed verbalizations was omitted four times, and one change (from flaps 1 to flaps up) was made without comment from either pilot. However, it was noted that aircraft altitude and location over the ground varied considerably at the initiation point of selected configuration change procedures (e.g., the Noise Abatement Departure Procedure), which were to be performed simultaneously.

Verbal indicators of the transfer of Exhaust Gas Temperature (EGT) Monitor and Transfer of Aircraft Control Procedures typify the quality of communications between specific pairs of crew members. In only two of the ten simulated flights does the flight engineer fail to advise the flying pilot when he can relinquish responsibility for monitoring EGT. However, in spite of obvious needs to effect the optional transfer of control procedure, two crews never use it and three crews execute incomplete double transfers. Only one crew uses more than two transfers (4) during the simulated flight.

It should be noted that verbal behaviors dictated by the aforementioned crew coordination procedures can reasonably be expected to enhance crew-coordination and flight safety. It should also be noted that non-compliance appears to depend more upon the operators involved than on the requirements of the procedures.

Errors and Procedural Compliance

The Ruffell Smith error counts have been modified and expanded so that every error is identified and individually related to an operator or group of operators. Those data are summarized in Table 2. The error categories coded by responsible operator are: pilot flying (PF), pilot not flying (PNF), captain, co-pilot, pilot team, flight engineer (FE), and entire crew. These categories cover all the errors recorded.

The next step was to investigate potential relationships between the enumerative error data and the enumerative procedure compliance data. Because of the limited sample size, relationships noted below should be taken as indications of fruitful directions for further research rather than as definitive results.

A set of fifteen dependent variable categories (error counts) was generated by creating various combinations of six of the categories noted in Table 2. A set of seven independent variables (five involving procedural compliance and two involving crew experience) was also generated as noted in Table 3. Stepwise multiple regression techniques were then used to identify the best models relating the independent (procedural) variables to each of the dependent (error) variables in turn. Results of that analysis, noting independent variables included and the maximum coefficients of determination, are shown in Table 4.

Dependent variables, which reflect errors by the flying pilot (PF, TPF, CPF), by the captain (CAP, TCAP, CCAP) and by the two pilots collectively and individually (PLTs), all have highly significant regression models in which pilot flying checklist commands (PFCK) and non-flying pilot callouts (PNFC) are the common independent variables. That is, pilot errors do appear to be related to those two classes of procedural non-compliance.

PROCEDURES SUMMARY

The study of procedural compliance by aircrews who participated in the Ruffell Smith experiment suggests the following observations:

1) Crew members face an impossible challenge in attempting to mentally catalog all of the standard operating procedures (SOP) published for them.

2) Routine non-compliance with an assortment of SOP's has been documented.

3) Forty-five percent of the enumerated crew errors involved two or more operators, which suggests that human redundancy by itself does not erradicate personnel error.

4) A statistical link appears to exist between operator errors and procedural compliance.

5) Full mission simulation offers new possibilities for studying aircrew behavior in a controlled, high fidelity, operational setting.

6) Altitude callouts, which duplicate functions performed by a machine, produced the highest frequency of non-compliant behavior, suggesting that they may need modification.

TABLE 2

Ruffell Smith's Errors[a] Attributed To Operators

		Experimental Run Number									
Operator(s)	Code	3	4	5	6[b]	8	10	12[b]	13[b]	14	15
Pilot Flying	PF	4	1	1	6	2	0	1	4	0	1
Captain	CAP	4	1	1	6	2	0	3	2	0	1
Pilot not flying	PNF	1	1	0	6	2	2	3	2	4	0
Copilot	COP	1	1	0	6	2	2	1	4	4	0
Pilot team	PTM	1	2	2	4	3	2	2	4	1	2
Flight engineer	FE	0	0	2	0	0	0	0	0	0	0
Entire crew	CRW	1	0	2	3	1	1	1	1	1	1

[a] Error categories and descriptions are individually related to operators in Appendix D of Schofield's dissertation (3).

[b] On runs 6, 12, and 13 the captain is the PNF and the copilot is the PF.

TABLE 3

Independent Variables

		Experimental Run Number									
Variable name	Code	3	4	5	6	8	10	12	13	14	15
PF checklist commands	PFCK	5	5	8	1	3	7	3	3	5	8
Pilot checklist announcements	PA	2	2	1	5	0	1	0	0	2	2
FE checklist announcements	FEA	3	3	3	2	0	2	1	0	1	0
PNF callouts	PNFC	11	13	13	9	15	16	15	13	15	14
Aircraft control transfers	TRAN	3	2	4	0	3	6	0	2	4	4
Crew members with more than 1000 hours in B-747	CREX	3	2	2	1	2	3	1	1	2	2
Pilots with more than 1000 hours in B-747	PEX	2	1	1	1	2	2	0	0	1	1

TABLE 4

Maximum Coefficients of Determination

Dependent variable[(1)]	Independent variables					
	One		Two		Three	
	Code	r^2	Codes	R^2	Codes	R^2
PF	PNFC	.749	PFCK, PNFC	.869	PFCK, FEA, PNFC	.912
CAP	PNFC	.651	PFCK, PNFC	.824	PNFC, TRAN, PEX	.878
PNF	PFCK	.579	PFCK, PA	.709	PFCK, PA, PNFC	.864
COP	PFCK	.513	PFCK, TRAN	.642	PFCK, PA, TRAN	.798
PTM	CREX	.446	CREX, PEX	.626	TRAN, CREX, PEX	.665
TPF	PNFC	.571	FEA, PNFC	.785	PFCK, FEA, TRAN	.848
TCAP	PFCK	.595	PFCK, PNFC	.825	PFCK, FEA, PNFC	.856
TPNF	PFCK	.692	PFCK, PA	.763	PFCK, PA, PNFC	.808
TCOP	PFCK	.567	PFCK, TRAN	.637	PFCK, TRAN, CREX	.789
CRW	PNFC	.317	PFCK, PNFC	.363	FEA, CREX, PEX	.550
CPF	PNFC	.595	PFCK, PNFC	.769	PNFC, CREX, PEX	.820
CCAP	PNFC	.569	PGCK, PMFC	.794	PNFC, CREX, PEX	.852
CPNF	PFCK	.606	PFCK, PA	.710	PFCK, PA, CREX	.743
CCOP	PFCK	.532	PFCK, PA	.624	PFCK, TRAN, CREX	.705
PLTS	PFCK	.597	PFCK, PNFC	.741	PNFC, CREX, PEX	.796

(1) Txxx refers to the pilot team plus member xxx.
Cxxx refers to the crew plus member xxx.

7) Lack of unitary leadership and internal coordination was most often observed when the captain was not flying the aircraft, suggesting a need to redefine flying co-pilot responsibilities.

REFERENCES

1. Rockwell, Thomas H. and Walter C. Giffin, *An Investigation Into Pilot and System Response To Critical In-Flight Events*. Final Report NAS 2-10047, National Aeronautics and Space Administration, Ames Research Center, Moffett Field, California, June, 1981.

2. Ruffell Smith, H. P., *A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance, and Decisions*. NASA TM-78482, Washington, D.C., January, 1979.

3. Schofield, Jeffrey E., "Aircrew Compliance with Standard Operating Procedures as a Component of Airline Safety", Doctoral Dissertation, The Ohio State University, 1980.

INSTRUCTIONAL DESIGN FOR AIRCREW JUDGMENT TRAINING

Dr. F. Brecke
Logicon, Inc.
San Diego, California

ABSTRACT

Aircrew training design has made significant progress during recent years. However, significant gaps exist both in design methodology and existing programs with respect to systematic training of higher level cognitive skills. Training in decision-making and judgment is currently haphazard at best. After a brief review of recent literature, the paper presents a conceptual model of judgment performance. The theoretical model is an extrapolation from Jensen (1977) and unites the variables of cognitive complexity, time availability, uncertainty, and stress in one coherent model. The model is used to examine current aircrew training and to develop new training strategies for improving judgment performance.

INTRODUCTION

The secret to being a good pilot is good judgment. Very few people familiar with the art of flying airplanes would disagree with that statement. However, the art of teaching judgment to a pilot is a secret. Very few people who deal in the magic arts of training design would quarrel with that statement.

Judgment is a part of the "RIGHT STUFF" of Thomas Wolfe's newest novel. A good student is supposed to come with a good foundation stock of that right stuff. During training, he might increase it slightly by mere association with his instructors who have plenty of it, of course. But the rest of it, the delta which will make him a full member of that exclusive elite of good pilots--that delta he will acquire by experience, by hours and hours of flying, by countless exposures to situations which require that rare quality of good judgment.

The great teacher then, of judgment is experience, and us ISD people who are so adept in teaching facts, concepts and rules and even such high level skills as problem-solving, grudgingly leave the field to Master Experience when it comes to instilling that most desirable of all skills, the skill of making good and reliable judgments. That this is not a desirable state of affairs, and that the guild of instructional designers in cooperation with the learned colleagues from the field of psychology should do something about it, goes without saying.

This paper presents a conceptual framework for thinking about the problem of judgment and a discussion of the applications of this conceptual framework to training design and research. This may sound like an ambitious endeavor, but if it provokes nothing but your passionate criticism, the cause of finding the instructional treatment for judgment may be advanced.

The most recent effort in the area of judgment training is a project sponsored by the Federal Aviation Administration (FAA). Phase I of this project resulted in a report by Jensen and Benel (1977) from the Aviation Research Laboratory at the University of Illinois. Their report was based on a literature review which, to quote the authors, "led to many studies related to pilot judgment in a peripheral way, but only to one which was directly related" (Thorpe, Martin, Edwards and Eddowes, 1976). The Jensen and Benel report is essentially the first broad-based attempt to answer the three questions: "What is judgment?", "Can judgment be trained?", and finally, "Can judgment be evaluated?".

The first question was answered with a definition which consisted of two components:

Judgment consists of:

1) A cognitive component which deals with establishment of alternative actions and the selection amongh them, and

2) An affective component or motivation which effects such selections among the alternatives.

In a later paper, Jensen expanded this definition to include a continuum from perceptual judgments to cognitive judgments (Figure 1).

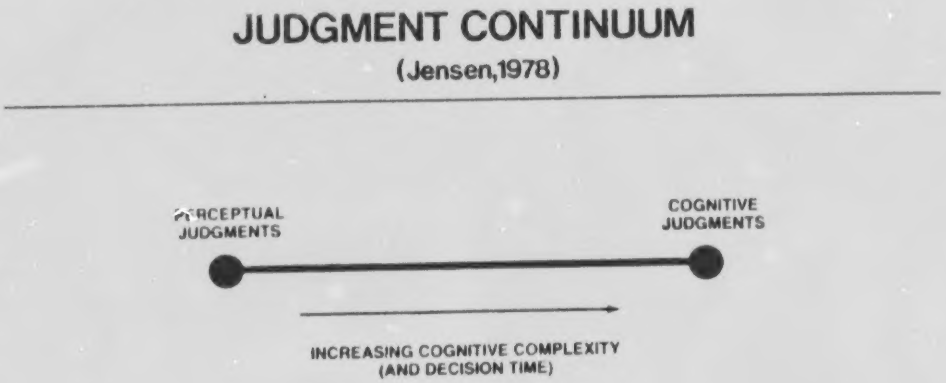


Figure 1. Judgment Continuum Based on Cognitive Complexity and Decision Time

According to Jensen, primarily perceptual judgment is associated with low cognitive complexity and little time available for decision making. At the other end of the continuum one finds more analytical forms of judgment, called cognitive judgment, characterized by high cognitive complexity and plenty of available decision-making time.

The question, "Can judgment be trained?" was answered positively. Judgment is trainable if Instructional Systems Design (ISD) methods are used for training design. The instructional designer, however, finds little operational advice which would enable him to "do his thing" in the area of judgment, except the recommendation to use situational training techniques.

The question dealing with the evaluation of judgment was also answered positively. Judgment performance can be evaluated if such an evaluation is based on criteria established by behavioral objectives and if the training situation is carefully structured. Again, however, Jensen and Benel offer no practical method of structuring the training situation.

Neither the recommendations for training, nor the recommendations for evaluation follow from the definition. The definition itself clearly is comprehensive enough to encompass all that might be called judgment, but it is not sufficiently precise to allow the logical deduction of training or evaluation methods.

In additon to the Jensen study, there is a fairly extensive body of literature which deals with aircrew decision making in a probabilistic environment. A good example of this literature is a study sponsored by the Naval Training Equipment Center analyzing requirements and prescribing a methodology for decision training in operational systems (Saleh, Leal, Lucaccini, Gardiner, Hopf-Weichel, 1978). Decision making, however, is only one, although the most complex one, of the cognitive activities to which the pilot must apply judgment. The recommendations from this body of literature, therefore, may not be applicable over the entire spectrum of the judgment phenomenon.

The training designer, if he is not to leave the field of judgment training to Master Experience, needs a definition of judgment which is both broad enough and precise enough. It must be broad enough to include the entire judgment spectrum and it must be precise enough to permit the deduction of design principles for training. In the following, an attempt is made at a redefinition of the concept of judgment and the major variables which influence judgment.

JUDGMENT MODEL

Definition

Fundamental to the understanding of the notion of judgment is the idea of uncertainty in the sense of lack of information. Judgment must be exercised when less-than-perfect information is available, i.e., under conditions of uncertainty. To express this in more colloquial terms: judgment is making more or less educated guesses if one does not know everything one should know in a given situation. For example, the pilot who cannot determine how far the wheels of his airplane are off the ground during the final phases of his landing approach, must exercise judgment in establishing the correct landing attitude. The same is true for a pilot in an air-combat situation, when faced with an opponent whose personal capabilities he does not know, whose weapons he cannot identify, and who flies an obviously modified aircraft of an otherwise familiar type. This pilot, too, is operating under conditions of uncertainty, i.e., under conditions of a relative lack of information, and must exercise judgment when he goes through the complicated decision-making process of determining his next maneuver, i.e., in deciding and planning how to best wipe his opponent out of the skies.

These two examples, which range from simple to complex cognitive tasks, illustrate another important point. This point is, that judgment is not a type of intellectual skill as identified, for example, in Gagné's hierarchy of intellectual skills or in Markle's, Bloom's or Merrill's taxonomies of objectives. Rather, judgment is exercised, and must be exercised, when performing any cognitive task, regardless of its taxonomic classification, when less-than-perfect information is available, i.e., whenever a cognitive task of any type must be performed under conditions of uncertainty.

Under conditions of zero uncertainty, with all necessary information available, the mode of cognitive operation in any cognitive task is not judgment, but a different mode of operation which one might simply call determination, since the outcome is essentially deterministic as with the solution of a mathematical problem. On the other hand, under very high or extremely high uncertainty conditions, the mode of operation might more appropriately be called intuition, a mode which might no longer be classified as a rational/cognitive mode of operation, as it borders on irrational phenomena such as the famous "sixth sense" or ESP.

The concept judgment, therefore, can be defined with three relevant attributes:

- It is a mode of cognitive operation which is delimited by strictly deterministic modes on the one side, and by intuition on the other, where the borderline between judgment and intuition is less sharp than that between the deterministic mode of operation and judgment.
- It occurs under conditions of uncertainty.
- It is applicable to any type of cognitive task, regardless of its taxonomic classification.

If this definition clarifies the concept of judgment, it does little to explain directly why different people faced with the same situation will operate with different degrees of success under conditions of uncertainty, i.e., will demonstrate anything from bad to good judgment. Given the relative paucity of the literature on the topic of judgment, it appears safe to postulate that the ability of a person to exercise judgment depends primarily on four factors or variables:

1. The difficulty of the judgment task, where the concept judgment task is defined as any cognitive task which must be performed under uncertainty conditions;
2. The available repertoire of relevant cognitive strategies, where cognitive strategies are defined in the sense of Gagné and Briggs (1974) as capabilities allowing a person "to manage the processes of attending, learning, remembering, and thinking."
3. The level of stress at the moment of tasking and the amount of stress generated by the task;

4. The *available repertoire of affective coping mechanisms* for dealing with stress.

This appears to be a quite plausible set of factors which can be subdivided into two subsets, the first dealing with cognitive factors and the second dealing with affective factors. Face validity of this set can easily be demonstrated with air combat. Whether or not a given pilot will demonstrate good judgment in an air combat situation, would certainly depend on his mental and emotional makeup and the perceived or objective danger and difficulty of the situation in which he finds himself. Physical stress and the physical condition of the pilot certainly influence judgment performance also. They are excluded here because the intent is to shed some light on the cognitive and affective aspects of judgment training only. Besides that, the physical conditioning part of pilot training really does not present any problems.

In the following now, the factor *difficulty of the judgment task* is defined with more precision and its relationship with *stress* is investigated.

Difficulty of the Judgment Task

One of the variables which determines the difficulty of a judgment task is the degree of *uncertainty*. The higher the degree of uncertainty, i.e., the less information is available in a given situation, the more difficult it will be to exercise judgment. Uncertainty and information are understood here in the sense of Shannon's Information Theory which defines information as a measure for the reduction of uncertainty resulting from the reception of a message and which establishes the bit as a measure of information. The uncertainty variable can thus be quantified in bits and the uncertainty of a situation is given by the amount of information in bits that is necessary for reducing undertainty to zero. In a binary situation, for example, the toss of a coin, the information "heads" has the value one bit and reduces uncertainty over the outcome of the coin toss to zero.

A second variable which determines the difficulty of the judgment task was identified by Jensen as *Cognitive Complexity*. One possible quantification of this variable is the number of discriminators and operators in an algorithm which describes the cognitive task. For example, the task of identifying a concept with three relevant attributes, can be described with an algorithm containing three discriminators and three operators. Saleh *et al*. (1978) have described several emergency decision tasks in algorithmic fashion, using decision trees.

The third variable is quite obviously the *Time Constraint*, i.e., the time available to perform the task. A task of a given complexity and uncertainty obviously becomes more difficult the less time is available to perform it.

For the sake of conceptual convenience, more than for reasons of theoretical necessity, these three variables can be arranged as the three axes of an orthogonal system of coordinates, or as three vectors originating from the same point. Difficulty of the judgment task is then simply defined as the resultant vector.

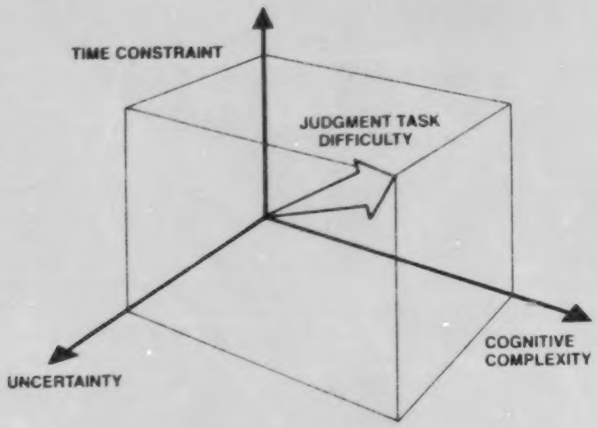


Figure 2. UCT Model: Judgment Task Difficulty as a Function of Uncertainty, Cognitive Complexity, and Time Constraint.

For the sake of rhetorical convenience, the three-axis model might be called the UCT model, the acronym being derived from the first letters of the three variables.

Stress

Before the application of the UCT model to training is discussed, the relationship of judgment task difficulty to stress or to affective variables must be analyzed. In order to characterize this relationship, an important distinction, the distinction between the flight problem and the background problem, must be made. Jensen and Benel make essentially the same distinction when they speak of rational and irrational pilot judgment. This distinction can best be demonstrated by an example comparing two flights. The purpose of Flight No 1. is a cross-country check ride for a private pilot license. One hundred miles before his destination, the pilot receives a weather report which tells him of an unanticipated deterioration of the weather between his present position and his destination with the destination itself still under VFR conditions. The pilot now has to exercise judgment in deciding whether to continue to his destination, to his preselected alternate, or to some other airport, and whether to fly through the weather, around the weather, or above the weather. In Flight No. 2, exactly the same situation exists. One hundred miles before his destination, the pilot receives the very same weather report. He flies the same aircraft, etc. However, the purpose of this flight is not a cross-country check ride, but an important business meeting, or a reunion with a lover, or a medical emergency. In Flight No. 1, the pilot is faced with solving a pure flight problem. He will exercise his best judgment, given the facts and uncertainties of the flight situation alone. In Flight No. 2, the same flight problem exists, but in addition to that, a second problem which can be called the background problem, is to be solved. The background problem and the flight problem obviously interact with each other. It can be hypothesized that if this interaction between these two problems produces conflict and/or cognitive dissonance (as defined by Festinger), a component of stress is introduced into the situation which can be considered additive to the stress already existing at the time the problem emerged.

Stress, however, arises not only from conflict created by the interaction between the flight and the background problem, but also from the flight problem itself. The difficult of the judgment task in the pure flight problem may also give rise to stress. The relationships between stress and task difficulty have been extensively investigated in human factors research, and by and large allow the conclusion that performance will initially rise with increasing stress up to a certain point, where it rapidly drops off to near zero level. Among flight instructors in military flight training, this phenomenon is well known as the so called "IQ Dump". There are probably very few pilots, especially in military aviation, who have not at one time or another experienced this phenomenon themselves, either in training or during operational flying. By the way, military aviation is emphasized here, not because it is assumed that the men (and nowadays women) who choose to become military pilots are especially likely to "dump" their IQs, but because military aviation is especially demanding.

Stress can thus be seen as a fourth vector in the model, adding an affective component to the definition of judgment task difficulty (Figure 3).

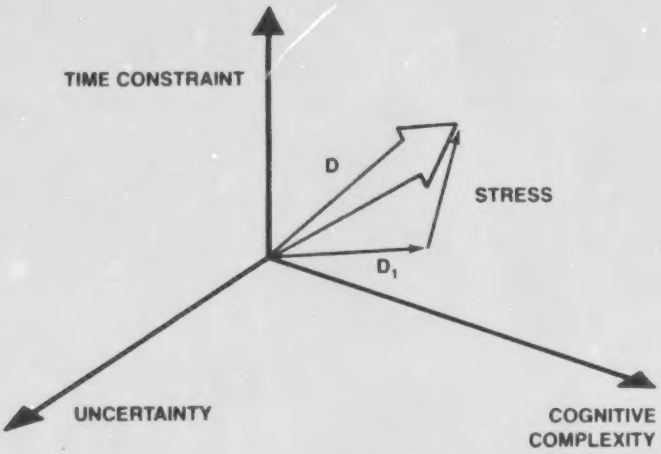


Figure 3. UCT Model: Stress as the Fourth Vector Determining Judgment Task Difficulty (D).

So far, then, two of the four factors likely to influence judgment performance have been discussed. The following conclusions seem warranted:

- Judgment task difficulty can be seen as the resultant vector of cognitive complexity, uncertainty, and the inverse of time availability.

- Stress will affect judgment performance in a non-linear fashion: positively up to an individual maximum, and negatively beyond that. The stress in a situation requiring judgment can be thought of as consisting of three components: the null-level stress, stress resulting from the difficulty of the judgment task itself, and stress resulting from the interaction of the flight problem and the background problem.

Types of Judgment Tasks

Before discussing the application of the model, one final distinction should be introduced. It is possible to distinguish between two types of judgment tasks. The first type, Type 1, occurs when the flow of some routine evolution is disturbed by some unpredicted, unforeseen occurrence (such as in an emergency or malfunction), or when something planned and predicted has not occurred (such as when the tanker aircraft does not show up at the planned rendezvous). Many such situations are covered by procedures, emergency and otherwise, but the potential variations in the situations are so innumerable that it would not only be inefficient, but virtually impossible, to invent a procedure for every eventuality.

The second type of judgment tasks are those tasks where the course of events is never routine, never predictable, always uncertain, and where judgment is therefore always required. This type of judgment task may be called a Type 2 task and air combat maneuvering is a perfect example for this class of judgment tasks.

APPLICATION OF THE MODEL

The two remaining factors, availability of relevant cognitive strategies and availability of relevant affective strategies for dealing with stress, will be discussed in the dual context of selection and training.

Selection

In the selection process for training in the Israeli Air Force, the applicants are subjected not only to paper-and-pencil tests and personality interviews, but also to situational tests which tax an applicant's intellectual resourcefulness and his ability to withstand psychological stress in practical, hands-on situations. It appears that these selection procedures are quite effective and that their effectiveness can be directly understood in view of the preceding discussion. Given the conceptual model above, it would appear possible to design carefully graduated situations varying in cognitive complexity, uncertainty, time availability, and stress that would yield valid predictive scores, when referenced to scores of a norm group of successful pilots.

Training

If the conceptual model presented above is applied to training, one can draw conclusions concerning the validity of current aircrew training methods and one can begin to define new training strategies specifically aimed at the development of good judgment performance under various degrees of psychological stress.

Current military aircrew training is characterized by an emphasis on correct completion of prescribed procedure, compliance with rules, and specific flight techniques. Judgment is a coveted quality and is usually evaluated under such headings as "headwork" or "airmanship", but explicit instruction in headwork or airmanship is almost totally absent.

The author has personally received this type of training in the U.S. Air Force and can attest to its inadequacy based on his rude awakening upon assignment to an operational squadron of the German Luftwaffe. A former F-14 instructor pilot describes current training practices in the Navy as follows:

> "The student is taught from day one, that he must use correct procedures. He is drilled for hours in these procedures and learns to accomplish tasks using the same steps each and every time. Use of checklists to cover normal, abnormal, and emergency procedures is mandatory.
>
> At the same time, he is presented with a myriad of rules to follow. Compliance with Course Rules, Air Traffic Control, NATOPS, SOP as well as Station, Wing, and NAVAIR instructions is taught.
>
> When reasonably proficient in applicable procedures and rules, the student is taught flying technique and is evaluated on "basic airwork". If the student has an opportunity to exercise judgment, he is evaluated under the heading of "headwork". The system is structured, however, so that virtually no opportunity exists, under normal circumstances, for a student to exercise his judgment. The student will be successful if he follows correct procedures and rules.
>
> Fleet replacement squadrons follow a similar approach to aircrew training. Flight techniques (including weapon system utilization) are more heavily emphasized than procedures or rules as a student is expected to master procedures and rules quickly with a high degree of transfer of knowledge from the Training Command.
>
> Graduates of the present system can be characterized as reasonably competent airmen, well-indoctrinated in applicable procedures and rules, whose decision-making abilities are relatively unknown.

> Once in the fleet, the fledgling aviator is expected to learn by example as he flies with an experienced flight leader and/or crewman. When a situation arises requiring exercise of judgment, he can normally count on being assisted by many experienced people, both in the air and on the ground.
>
> Problems arise in this system when the inexperienced aviator is cut off from the decision-making assistance of others. The student has very little to fall back on, except his procedures and rules. Unfortunately, there are not procedures to cover every problem that may arise.
>
> The aviator will probably decide on an acceptable course of action based on his knowledge and the decision-making skills and judgment that he developed prior to Navy flight training. . ."

When this state of affairs is examined in light of the model and the concepts presented previously, it is obvious that training in Type 1 judgment tasks is totally absent or accidental in two respects: it occurs only by accident, and when it occurs in the air, it is likely to cause an accident. Training in Type 2 judgment tasks exists only in the form of feedback, such as in the ACMR facilities when an air combat maneuvering sortie is replayed in three dimensions and each move and countermove is discussed post-facto.

Is explicit, systematic training of judgment really all that important? After all, the training pipelines do produce aviators who are able to function successfully in peacetime as well as in combat. The counter question here is: "At what cost?" It is well known, from military as well as from civilian accident statistics, that the overwhelming majority of aircraft accidents is attributable to pilot error and that most of these are simply errors in judgment. In other words: waiting for experience to teach you enough judgment can kill you.

Further insight into the importance of explicit and systematic judgment training throughout the pilot's career is gained when comparing jobs which are highly proceduralized by their very nature, such as the job of a bank clerk with the job of a military combat pilot. Figure 4 illustrates the task universes of these two jobs. They are arbitrarily limited to intellectual tasks using Gagne's taxonomy.

This illustration, by the way, is not based on actual data, but rather on sound judgment. The intent is to illustrate a principle rather than to provide exact quantitative data. The fact that the two task universes are different in overall size is irrelevant. What is relevant, however, is the fact that the task universe of the pilot contains much higher proportions of judgment tasks, i.e., incidents where tasks must be performed under conditions of uncertainty. There is little doubt in a bank clerk's job, once he has learned his rules and procedures. That is not the case in a pilot's job. When viewed in this light, current training practices essentially provide the pilot with bank clerk training. This is hardly in accordance with the much touted maxim "Train Like You Fight".

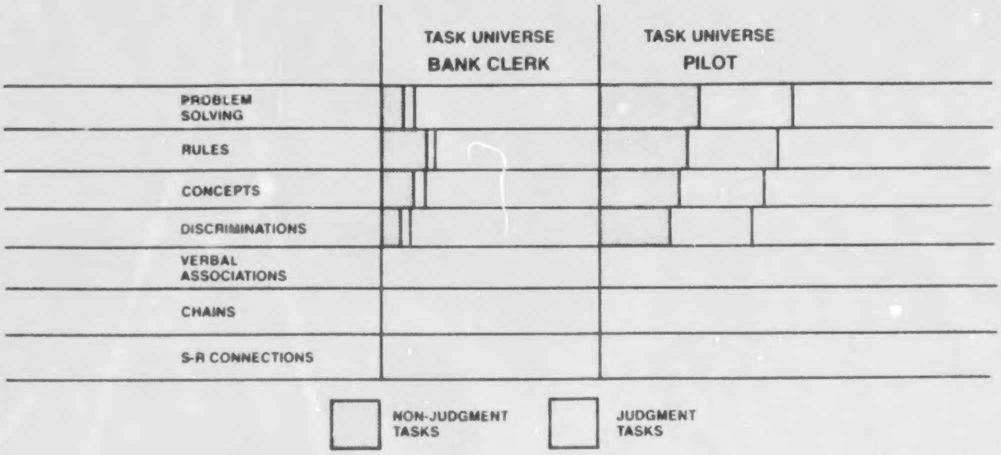


Figure 4. Proportions of Judgment vs. Non-Judgment Tasks in 2 Task Universes.

If a fighter pilot, or any other military pilot for that matter, must have one characteristic, one trait that is absolutely essential in the performance of his duties, that trait is probably initiative (besides judgment). However, trainees who are continuously given directive commentary and who are continually drilled in rigid rules and procedures will hardly develop initiative, but much more likely will abdicate responsibility and await orders. Training in judgment on the other hand, i.e., training under conditions of uncertainty, especially if such training follows an orderly step-by-step course, systematically develops a trainee's self-reliance and with that, his initiative.

One more point should be made regarding the importance of including judgment training in aircrew training curricula. The current rigid regimentation and proceduralization of training frequently leads to actual fear of departing from established procedures. Not all emergency situations fit precisely the case for which a given procedure is designed. As a consequence, the rigidly trained aviator who is unable to and fearful of bending the procedure might end up bending the airplane.

Given these arguments, the next logical question is: "If judgment training is so important, then why is it not included in the average aircrew training curriculum?" Two reasons come to mind. First of all, it is much easier to train overt, observable behavior. This state of affairs has not changed with the advent of ISD. Task analyses and objectives hierarchies for aircrew training rarely contain the covert, higher level cognitive tasks required of the combat pilot.

The second reason is the well known phenomenon where each accident spawns another rule or procedure to be included in the flight manual. The attempt to reduce costly errors due to bad judgment by regulation and proceduralization is like trying to cure the problem by treating the symptoms instead of the causes. In the words of one instructor pilot: "The exercise of judgment is inversely proportional to the amount of regulation.

The amount of regulation is usually proportional to the cost of judgment errors." Or, to say it somewhat differently: The attempt to remedy judgment errors by proceduralization starts a vicious circle which only leads to more errors in judgment. The only cure is to train judgment systematically.

New Training Strategies

The model presented in this paper offers a methodology by which such systematic training can be conceptualized, designed, and implemented.

First of all, judgment training must enable the trainee to distinguish the flight problem from the background problem. As mentioned before, background problems may include such things as medical emergencies, business trips, and lovers. However, they can be considerably more subtle than that. One background problem tha the military pilot is especially prone to encounter is excessive machismo, brought about by peer group pressure. This brings to mind a famous Royal Air Force proverb: "There are old pilots and there are bold pilots, but there are no old bold pilots." The "can do" attitude that is especially prevalent in today's Navy, quite likely fosters a scarf-flying, devil-may-care attitude rather than good and sound judgment. Given a behavioral objective such as "The student will identify the background problem, given verbal descriptions of flight scenarios, correctly in x out of y cases," it should not be too difficult to design appropriate instruction. Such instruction would generate in the student an awareness of how the background problem may bias or interfere with his judgment in solving the flight problem.

Secondly, the three-pronged model of uncertainty, cognitive complexity and time availability, allows systematic design of scenarios which are carefully graduated along these three variables. If the student is subjected to successively more difficult judgment problems under gradually increasing psychological stress, and if all such training sequences are administered on the basis of reaching criterion performance before progressing to the next higher level of task difficulty, it would appear reasonable to expect a high degree of training effectiveness.

For example, in any given area of flight instruction, be it familiarization, navigation, air-to-ground tactics, air-to-air tactics, etc., the model allows the design of a variety of possible systematic training strategies which lead from the least difficult to the most difficult of judgment problems. One such possible strategy is presented in Figure 5 in the form of a series of histograms which show relative quantities of the four variables of uncertainty, cognitive complexity, time availability and stress.

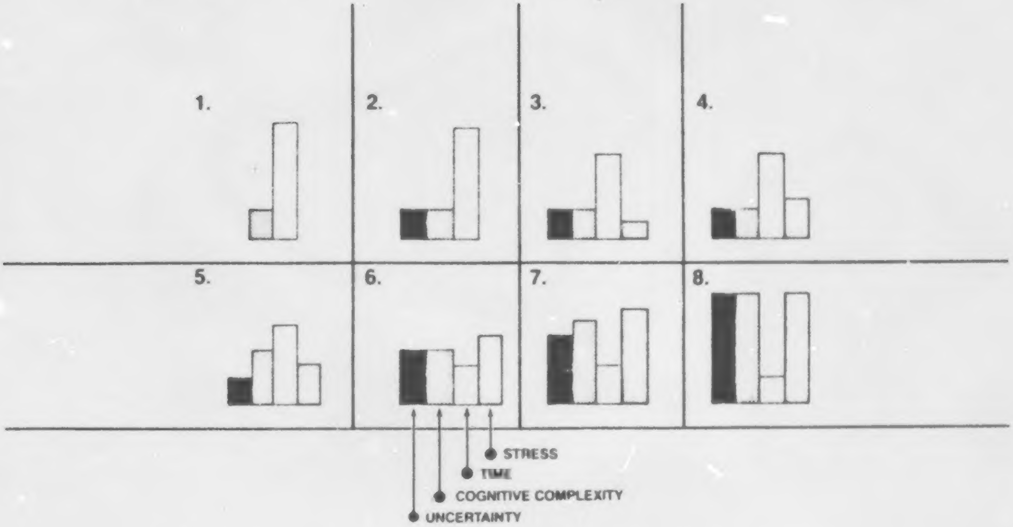


Figure 5. Histogram Representation of an 8-Step Strategy for Judgment Training.

As the figure shows, the student is initially presented with scenarios or situations of low cognitive complexity and ample time availability. As he becomes proficient, an element of uncertainty is introduced while keeping complexity and time constant. Gradually, stress is introduced, the time available is decreased, and the levels of uncertainty and cognitive complexity are built up until the student can finally demonstrate acceptable judgment performance over the entire spectrum of the particular subject matter domain being trained.

The gradual build-up of stress deserves special attention. Classroom instruction, i.e., the academic environment, presents very little natural stress. Stress is somewhat increased when the student transitions from the classroom to the simulator environment, since he is there operating under real-time conditions and under the influence of much more sensory stimulation than in the classroom. The transition from the simulator environment to the flight environment, however, represents a quantum jump in stress. The reason for this is quite simply, fear of death. The simulator cannot crash. It is, therefore, important that the student is accommodated to these higher levels of stress in the preceding simulator and academic environments. It is equally important not to overdo this principle so that the student does not develop fear of flying.

A second note of caution should be introduced here. Current training essentially remains in the deterministic plane (Figure 6). The training situations which are administered are usually canned, and have very little built-in uncertainty. As mentioned previously, the student learns innumerable rules and procedures. On the face of it, it would seem quite reasonable to retain this sort of training and to proceed along the uncertainty axis only after the student has acquired a firm basis in the deterministic plane. One could, for example, use in tactical decision-making, highly complex algorithms that the student works through in a

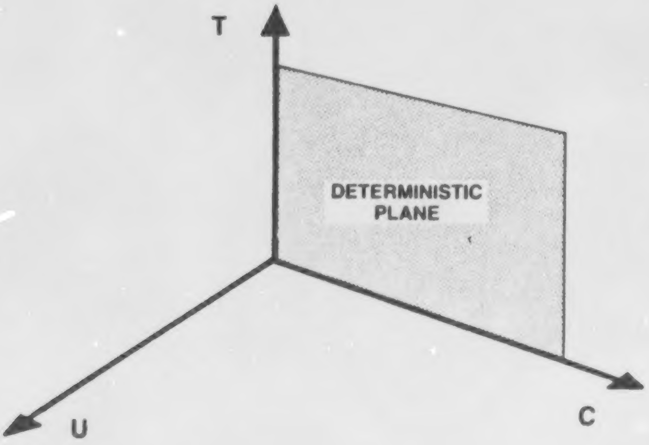


Figure 6. UCT Model: The Deterministic Plane

classroom environment. It is doubtful whether training in such deterministic modes of operation has a high degree of transfer to judgment performance which is an essentially different mode of cognitive operation, i.e., a probabilistic mode. Flying will never be totally predictable, and any attempts to transform the pilot into a high-speed automaton or computer are doomed to failure. The development of cognitive strategies that are likely to be successful in an uncertain environment is probably fostered more effectively and efficiently, if uncertainty is introduced all along, i.e., if the student is led along a path which proceeds step-wise along each of the axes.

Research

The preceding discussion of training design strategies based on the UCT model immediately leads to numerous questions which can only be solved by experimentation. It is suggested that the UCT model is an exceptionally productive conceptual vehicle for the design of research programs because of its face validity on the one hand, and because of the many questions it poses, on the other hand. It appears suitable not only as a vehicle for investigating instructional strategies for judgment training, but also for the investigation of the judgment phenomenon itself. The aircrew training community, however, can hardly afford to wait until the results from such a research program, which would take years to put into motion and years to conduct, are available. Such waiting would only perpetuate the depressing accident statistics on pilot error and endanger the combat readiness of our Armed Forces.

Beginnings

The aircrew training community has, in fact, already begun to implement new forms of training which are likely to foster the development of good judgment. For example, the F-15 program features situational emergency training (SAT). In this program, the old style boldface emergency procedure training is totally absent. The student learns only three rules:

1. maintain aircraft control, 2. analyze the situation and take proper action, and 3. land as soon as practicable. The student is presented with carefully structured scenarios, both in the classroom and the cockpit procedures trainer, and is consistently encouraged to use his judgment in finding the proper action for this problem.

Another example is the air combat training syllabus for the F-14 aircrews, developed by Veda Incorporated. In this syllabus, the pilots are first trained in the classroom in the general principles of maneuvering an F-14 in an air combat situation. They then proceed to the ACMR debriefing facility which is used in a pro-active manner, rather than in the debriefing/feedback mode. The student is there presented with a graphic image of a typical engagement, and on a second screen, with such data as airspeed, altitude, etc., for both the fighter and the opponent. The dynamic graphic presentation is frozen at crucial points, and the students are asked to assess the situation either from the standpoint of the fighter or the opponent, and to exercise their judgment in deciding on their next move. It is quite possible (however, not presently implemented) to create fairly high levels of stress by increasing the pressure for quick answers and by punitive comments for wrong or slow answers. The students then transfer to the simulator environment where they fly the same types of engagements. The simulator of course also offers the advantage of decision time manipulation by the freeze capability. Stress levels again can be manipulated by the manner in which the instructor reacts to the student's actions. The student finally proceeds to the flight environment, well prepared for the split-second judgments he must make under stress. He no longer finds himself thrust suddenly into an environment which, compared to his preceding instruction, shows simultaneously enormously increased levels of uncertainty, complexity and stress, and much less available time. The phenomenon of the IQ dump will hopefully occur with less frequency, but it is as yet too early to tell.

SUMMARY

In his recently published book, "The Third World War", General Sir John Hackett advances the hypothesis that said world war will be won by --among other things--the higher degree of independence in junior leaders in the tactical area. Such independence requires judgment, good judgment under stress. There can be little doubt about the importance of systematic judgment training not only for aircrews, but throughout the Armed Forces.

This paper suggested a conceptual model and a design approach for systematic judgment training. The ideas presented here, albeit not tested by either experiment or experience, are simple and operational enough to be put into practice immediately.

REFERENCES

Gagne, R.M., and Briggs, L.J., Principles of Instructional Design. New York: Holt, Rinehart, and Winston, Inc., 1974.

Hackett, Sir John, The Third World War, August 1985. New York, N.Y.: Macmillan Publishing Co. Inc., 1978.

Jensen, R.S., and Benel, R.A., Judgment Evaluation and Instruction in Civil Pilot Training. Report No. FAA-RD-78-24 by Aviation Research Laboratory, Univ. of Illinois, December 1977.

Jensen, R.S., Pilot Judgment: Training and Evaluation. Proceedings of the 11th NTEC-Industry Conference, Nov. 1978, pp. 71-83.

Saleh, J., Leal, A., Lucaccini, L., Gardiner, P., and Hopf-Weichel, R., Analysis of Requirements and Methodology for Decision Training in Operational Systems. NAVTREQUIPCEN TECH REPORT No. 77-C-0005-1, February 1978.

Thorpe, J.A., Martin, E.L., Edwards, B.J., and Eddows, E.E., Situational Emergency Training: F-15 Emergency Procedures Training Program. William Air Force Base, Arizona, Technical Report No. AFHRL-TR-76-47 (1), June 1976.

ABOUT THE AUTHOR

Dr. Fritz H. Brecke is a member of the technical staff of Logicon, Inc. Tactical and Training Systems Division, San Diego, California.

AIRMANSHIP: AN INTRODUCTION

A sage's question is half the answer (a Jewish proverb).

Amir Mané
Cognitive Psychophysiology Laboratory
University of Illinois

Introduction

Airmanship is defined in the dictionary as "skill in piloting or navigating airplanes" (Webster, 1977). This article does not deal with the meaning of the word as such. Airmanship (or more accurately, its Hebrew equivalent - "Aviraut") has a different meaning when used by the Israeli Air Force. The term is used widely to describe a pilot's proficiency and a cadet's potential to succeed. It is generally believed that Airmanship is one of the more important aspects of flight skill, although it lacks a precise definition. Whenever decisions are being made regarding the potential of a flight cadet to become a fighter pilot, his level of Airmanship is considered as a major factor in the prediction of success. Flight instructors, discussing the manifestation of this phenomena, tend to agree in their evaluation of any particular incident of Airmanship. Yet, there is no mutual agreement on what the term describes nor on its objective definition. Instructors regard it as an undefinable, holistic, criterion of a pilot's quality of performance. A pilot's Airmanship reflects an overall evaluation of his understanding, of the flight situation as manifested in his decision-making.

In a systematic effort to clarify what is conveyed by this term I have discussed the topic in various forums of pilots. I have interviewed experienced flight instructors and squadron commanders, and solicited incidents which exemplify Airmanship. My goal was to define and describe this trait which is regarded as the essence of the flight skill. I will state briefly the results of this work and then describe it in detail. From a study of the examples and attributes, a distinction between two uses of the term emerged. Airmanship is used to describe a level of performance, but, it is more often used to describe a personality trait. As performance, it describes excellence in the decision-making and judgment of the pilot. As a trait, it is the ability and the tendency to define the relevant aspects of an aerial problem or opportunity.

Some Examples of Airmanship

All the examples were actually given by flight instructors. Some of them are not related at all to aviation and still serve as good examples of Airmanship.

Airmanship means not asking the control tower for permission to do something, when it is obvious that they would not grant the request (e.g., permission to enter the runway while there is another airplane in the "final").

Airmanship means to be able to drive your car on a very long street without stopping even once at the lights.

If you come to the base through a point where the height of reporting is 6,000 feet and the approach control tells you to report at 7,000 feet, good Airmanship means to start looking for the other airplane which is most probably ahead of you somewhere.

After shopping, to think ahead of time about which cashier's line will move faster is good Airmanship.

When exercising right above your base, (and according to the approach procedure, you would have to make a long detour), good Airmanship is to get permission to exercise emergency landing and in this way to get straight down to the base.

Airmanship as Performance

The first use of Airmanship is to describe performance. In order to understand this use we will have to scrutinize the nature of flight skill. When describing this complex skill it is useful to distinguish among three different components. A pilot must master three subskills: "what to see," "what to do" and "how to do." A simple demonstration will clarify this abstraction. A pilot on the runway, ready to take off, knows "what to see," he knows which of all the plane's instruments is relevant to the task at hand. He sees the wind direction and velocity by looking at the wind bag. When starting the takeoff he sees the runway, all the external cues that help him stay in the center of the runway and determine his speed. The pilot knows "what to do"; what he should do to keep the airplane from turning into the wind, where to direct his plane if he finds himself off center, what he should do if the fire warning turns on in the beginning of the takeoff and what to do if it turns on when the speed is high. The pilot knows "how to do," how to manipulate the tail rudder and keep the plane from turning from one side to the other, how to push the throttle without causing the engine to ramble, how to take the airplane off the ground without stalling it.

With this description of the nature of flight skills, we can eliminate the third element from our discussion. Airmanship does not reflect the motor component of the skill. None of the examples or descriptions that were collected, refered to the motor component, which, important as it may be, does not constitute any part of Airmanship. The example of driving fast in a city represents a complexity of consideration that the driver has to take into account in order to be able to go through the long street without having to stop. This example seems to reflect the performance component, but the essence of this excellence in performance is in the cognitive skills. The driver has to consider the features of the road, he must be familiar with the synchronization of the stop lights, judge constantly the behavior of other drivers on the road, and, probably, watch for the police as well. His actual behavior is compared against an ultimate criterion of excellence. Defining this criterion will be far beyond the scope of this paper. The "art of flying" is a phrase which attempts to grasp this subtle concept. For our purpose we will settle for less, we will conclude that it is this criterion which underlies the "integration of the pilot in the aerial situation." An inevitable question which stems from this description is "what is it that makes one pilot better than the other in those abilities?" "What is the essence of this human trait"?

Airmanship as a Personality Trait

The first and second components of flight skill aforementioned, seem to be connected. What to see and what to do can hardly be separated. The importance of the different aspects of the flight situation keeps changing all the time. Knowing where to look is dependent on knowing what you want to do. Vibrations in the fuel indicator are of great importance when you see them a short time after take-off. But, they are almost meaningless to the task itself when they appear in the final approach. The compass provides almost no information for the pilot when the dog fight is at peak, but is valuable when trying to make your way back to the base. The decision as to which of all the flight dimensions requires attention is highly dependent upon the purpose of the pilot and the action planned by him. On the other hand, it is the information gathered that enables you to decide what to do. It is in this connecting point between the knowledge of what to do, and the knowledge of what to see, that people differ. This is the essence of the behavior described by Airmanship. Airmanship is *the ability and the tendency to define the relevant aspects of an aerial problem or opportunity*.

During his period of instruction, the flight cadet is required to give the answer to a variety of well defined problems. An example of one, would be navigating the airplane. After finding his location he goes through a detailed computation that results in a quantitative adjustment in the flight direction and in the engine setting. This is an example of a well defined problem. Another example would be a simulated emergency landing where the instructor tells the cadet to assume that his engines do not function any more. Again, the cadet has to go through a clear set of checks and procedures. In real life we more often encounter ill defined problems. The pilot faces an infinite amount of information that keeps changing every second. Information that comes from the instruments of the plane, visual cues from the outside, vibrations, smells and sounds, messages on the radio and the like. He also knows many facts regarding his goal and situation. The first and most crucial step toward solving the problem is defining the relevant aspects of the problem. The flight cadet who is asked to exercise an emergency landing has to define first the relevant aspects of the landing situation. Teachers, as a common practice, help students by further defining the problem, alluding to its relevant attributes. As an example, consider the flight instructor, who says: "So, where will you try to land if your two engines are down, and you are too far from your base, and right below us is a wide highway which is used infrequently?" By rephrasing the question, the instructor defines the problem further and by doing so, leads his student to the desired solution.

The ability to define the variables which are relevant to a problem depends on background knowledge, but given common background knowledge, cadets will differ in their ability to explicate from that knowledge. The personality trait named Airmanship includes two different components: "ability" and "tendency" to exercise that ability. The individual differences appear when ability is measured, but they are much more salient when the tendency is considered. Assume two people in a car approaching a stop light in a multi-lane road. The driver stays in his lane and stops behind three heavy trucks. He does so while two other lanes which lead to the very same direction are empty. When approaching the second stop light his companion asks him: "What would be the best lane for you to take?" "No problem", would answer our driver, and will turn to the empty lane this time. The driver has the ability to define the relevant aspects of the situation but apparently, lacks the tendency to do so. Let us go back to the pilot who is required to be at the point mentioned earlier at 7,000 feet instead of 5,000 feet. If he will try to ascertain why he was asked to do so, chances are that he will easily deduce the reason. The difficulty lies in asking the question. The pilot who exercises right above the base does not face a problem at all. He can fly according to the normal flight procedure without any difficulty. Yet, he has an opportunity, which, if he is able to define it, could help him achieve more during the flight lesson. The ability to see where the aerial opportunity exists is a manifestation of good Airmanship.

One attribute of Airmanship was emphasized by a flight instructor who said: "Airmanship is not a process that takes time. It is not that you think and think and then come to a solution. You simply see it, in no time at all". Very often in popular literature the pilot is described as someone who thinks very fast about many possible solutions and then creatively comes up with an optimal ingenious solution to a problem. Another instructor attacked this description. "A pilot is hardly ever required to really think, and find a solution. Flight is like a flowchart, you always know what to do in a given situation. The problem is just to know what the situation is." This once again shifts the emphasis from solving the problem to defining it.

Allow me to discuss one more example that serves to emphasize the difference between solving problems and Airmanship. When learning to fly the landing pattern the cadet is told what should he do if he finds himself drifting away from the appropriate location of the pattern. The pilot ought to make a small correction in his heading and gradually get close to the right pattern. A pilot who descends into the landing pattern and finds himself above the runway may attempt to solve the problem by the aforementioned method. Thus he will gradually approach the right pattern. Good Airmanship, in this case, would be to diverge from the regulations and change your direction abruptly. Being above the runway means that you are in danger of head-on collision with other airplanes which get into the pattern. Naturally, it is best to terminate this situation as soon as possible. The nature of the problem solving described is different than putting the variables of decision into a hypothetical equation. The good pilot immediately sees the implications of being above the runway. Knowing this, there is no question regarding what he should do. For the good airman, reality is represented not as seperated blocks of information but rather as factors which have implications. Seeing the implications of the various components of reality is what Airmanship is all about.

Performance and Trait

As in many other cases, it is impossible to assess a trait directly. It is performance - the behavior, which serves as our data. A correct interpretation of performance is essential to the evaluation of the trait.

This interpretation is not straightforward at all. A well trained pilot with thousands of flight hours has been in many aerial situations. His experience makes the relevant aspects of a common state of affair very salient. He does not have to deduce implications of an event, because, experiencing it many times in the past, he knows and remembers well the consequences of the particular situation. This distinction is even more prominent while interpreting the behavior of the inexperienced pilot. Airmanship is most salient when encountering a new situation. A cadet, who has been with one instructor in an unusual aerial situation, will be mistakenly credited by another instructor, for very good Airmanship when replicating the solution he saw the day before. For a novice pilot, it is a sign of good Airmanship to suggest a possible short cut in the approach. For an experienced pilot this would be a matter of routine, and would reflect the ability to use his experience, ability which is different from the one discussed in this paper. Clearly, the implication of this is that there is no clear-cut way to distinguish one from the other. When coming to evaluate a pilot's Airmanship, his past experience should be taken into account.

Closing Remarks

Airmanship is considered an essential component in the prediction of pilots' effectiveness in combat. Yet it has always been a vague term. The question "what is Airmanship" used to represent the search for an answer to a question that had no answers. In this paper, I briefly discussed the equivocality of the term as performance and trait and the two meanings of the trait as "ability" and "tendency." This conceptualization is not a hypothetical concept which attempts to serve as a psychological model but rather a concept which is used among pilots but was never explained. Many attempts were made to describe the pilot's decision-making process. Models were constructed, and flowcharts were designed. In this article, I have tried to describe the essence of the cognitive aspects of flight skill from the pilot's point of view. My hope is that this abstraction can lead to a better understanding of the process of judgment and decision-making in flight. This in turn should lead to better training methods and to construction of a selection test which will predict the success of the cadet based on this trait.

DEVELOPING A CIVIL AVIATION PILOT JUDGMENT TRAINING AND EVALUATION MANUAL

Jerome Berlin, Ph.D.
Charles Holmes, Ph.D.
Embry-Riddle Aeronautical University

INTRODUCTION

PURPOSE

The purpose of the program is to improve behaviors associated with judgments and to plan an experimental method to test it. The results are contained in 3 volumes. Volume I contains the concepts upon which the program is based and suggested evaluation methodology. Volume II, the Student Manual, contains the instructional material in judgment training and companion worksheets by which the student may determine his or her progress. Volume III, the Instructor Manual, provides the instructor with a systematic approach to administering the judgment training contained in the student manual and evaluating the performance of the students.

The overall approach used in the program addresses judgment behavior of pilots, uses self-assessment to determine poor judgment tendencies, role modeling of the instructor, and accepted principles of learning. We agreed in principle with the definition of judgment offered by Drs. Jensen and Benel. (Reference) However, we felt that the definition should be a little more specific as to the process, and the definition should be operationally based. Pilot judgment was therefore defined as follows for the purpose of this program only:

> "Pilot judgment is the mental process by which the pilot recognizes, analyzes, and evaluates information regarding himself, the aircraft, and the outside environment. The final step in the process is the making of a decision pertaining to the operation of the aircraft."

STUDENT MANUAL

The student manual contains 20 lessons which are divided into three units. Unit 1 of the student manual presents concepts and terms which are used throughout the judgment training course. These terms and concepts have been especially designed to lead the student into modified patterns of thinking which will ultimately produce better judgment. The terminology also gives students and flight instructors the most concise, objective means possible of discussing the pilot behavior in judgment related situations.

Four sets of new terms and concepts are presented in Unit I. The first of these sets introduces the student to Three Subject Areas relevant to pilot judgment: Pilot (P), the Aircraft (A), and the Environment (E). Conventional flight training deals with subject areas A and E only. The judgment increments emphasize the pilot's need to know more about the area P, the pilot; and how the pilot interacts with the aircraft, PA; and the flight environment, PE; and the combination PAE.

A second set of terms is called the six action ways. Nearly 600 NTSB accident briefs were examined to determine how pilots carry out the actions resulting from their decisions. It became obvious that pilot implemented poor judgment decisions in six ways. These decision situations are verb phrases called the Six Action Ways. The six possible actions ways are:

DO = The pilot did something which he should not have done.
NO DO = The pilot did not do something which he should have done.
UNDER DO = The pilot did not do enough when he should have done more.
OVER DO = The pilot did too much when he should have done less.
EARLY DO = The pilot reacted too early when he should have waited.
LATE DO = The pilot reacted too late when he should have reacted sooner.

The repetitive use of these terms or action ways is designed to effectively and positively identify the actual, erroneous response simultaneously with the desired response.

The third new set of concepts presented is called the Poor Judgment Behavior Chain. Research into accidents shows that once a poor judgment (PJ) is made, there almost always follows a sequence of poor judgments. As the chain of poor judgments grows, it is automatic that the number of alternatives diminishes very rapidly. If this sequence or chain is broken early in the situation, the pilot may have more alternatives for successful recovery. In the judgment training material, the phenomenon where one poor judgment leads to another is referred to as the PJ Behavior Chain. This section teaches the student about some of the chain mechanisms and what must be done to break the chain. The judgment program uses elementary behavior training to trigger within the pilot a new response pattern effectively breaking the PJ sequence chain.

With the fourth set of concepts, the student is taught to understand and apply what are identified as the Three Mental Processes of Safe Flight. The first of the three processes is called "Automatic Reaction" (AR). Two general categories of automatic reactions are taught. One involves the flight skills having to do with maintaining positive, on-going control of the aircraft. The other concerns those learned responses to unusual or emergency situations.

Problem Resolving (PR)is the mode of thinking that helps a pilot overcome undesirable situations by means of a systematic process.

Repeated Reviewing (RR) is the mode of thinking that allows the pilot to be continuously aware of all the factors (pilot/aircraft/environment) that affect safe flight.

Unit II of the Student Manual contains the behavioral aspects of the training. This unit is designed to adjust, or to redirect, a pilot's tendencies in such a way as to promote the consistent use of good judgment. The first approach addresses the pilot's hazardous ways of thinking, his hazardous attitudes. Five hazardous thoughts for pilots are identified, and an exercise for self-assessment of hazardous thought patterns is provided.

Since little prior research was found in which such thought patterns were described, it was necessary to consult experts to obtain informed judgments of the nature of such hazardous thought patterns, and to choose a descriptive name for the thoughts.

1. Anti-Authority. This is the thought patterns found in people who resent the control of their actions by any outside authority. The general thought is, "Do not tell me! No one can tell me what to do."

2. Impulsivity. This is the thought pattern found in people who when facing a moment of decision, feel that they must do something, anything, and do it quickly. This thought is characterized in the student manual as, "Do something - quickly!"

3. Invulnerability. This is the thought pattern of people who feel that nothing disastrous could happen to them, personally. The thought is characterized in the student manual by the statement, "It won't happen to me!"

4. Macho. This is the thought pattern of people who are always trying to prove that they are better than others. they prove themselves by taking risks and try to impress others by acting dangerously. The thought is characterized in the student manual by the statement, "I can do it!"

5. Outer Control. People who have this thought pattern feel that they can do very little, if anything, to control what happens. When things go well, it is attributed to good luck. When things go badly, it is attributed to bad luck, or it is generally the fault of someone else. This thought is characterized in the student manual by the question, "What's the use?"

It should be emphasized that the selected hazardous thoughts are preliminary only. Additional research and validation of the judgment training program is needed in order to fully develop the hazardous thought concept.

Next students are taught to identify and understand the five hazardous thoughts. A lesson is devoted to each particular thought. Next is a lesson which specifies substitute thoughts, called "antidotes", for the five hazardous thoughts.

Unit III, Applications, contains written lessons to relate the concepts of Units I and II to actual flight situations. No new flight or judgment information is taught. This unit is made up of numerous exercises centered about scenarios and case histories of pilots carrying out flight activities.

The scenarios and histories are taken from reports of actual accidents and incidents reported within the last five years. The unit is intended to make judgment training seem real; to make it come alive for the student. A sense of relevancy, of personal involvement, is essential for new behavioral learning to take place.

INSTRUCTOR MANUAL

The Instructor's Manual outlines for the instructor the material contained in the Student Manual, explains how the instructor is to present the material to the students, and provides guidance on how to resolve student difficulties regarding the course. In addition to describing how the instructor should present the program, the instructor manual contains material designed to enhance the influence of the instructor as a role model for the student. Finally, the instructor manual contains two sets of exercises for the instructor to conduct during flight training activities. These exercises are designed to further develop and focus the student pilot's judgment-making abilities as well as to reinforce the conceptual and behavioral aspects of judgment training. These exercises are generally completed in conjunction with other flight training activities. Thus, the increase in logged flight time for the student to complete the judgment training is minimal.

CONCLUSION AND RECOMMENDATIONS

It was the intent of this research to produce an experimentally derived training program. As of yet, no construction testing was done to verify either the appropriateness or the effectiveness of the training increments making up the program. Only through such testing can we know which materials and sequences are optimal. It may be shown by empirical testing that we have constructed too much material, that it needs a different order of presentation, or that different materials need to be developed. We, therefore, recommend that a validation of the program be conducted under controlled conditions prior to the time that the program is field tested. An additional advantage to conducting controlled validation is that optimum packaging guidelines can be developed as a corollary of the validation. Finally, the program should be assessed for suitability in other areas of aviation and its adaptability in other endeavors such as industry and driver training.

REFERENCE

Jensen, R. S. and Benel, R. A. *Judgment Evaluation and Instruction in Civil Pilot Training*. Washington, D. C.: U. S. Department of Transportation, Federal Aviation Administration, December 1977. (NTIS, Report No. FAA-RD-78-24)

PLANNING BEHAVIOR OF PILOTS IN ABNORMAL AND EMERGENCY SITUATIONS

Gunnar Johannsen

Research Institute for Human Engineering
Wachtberg-Werthoven, F. R. Germany

William B. Rouse

University of Illinois
Urbana, Illinois 61801

Two studies of the nature of planning behavior of pilots will be discussed. An experimental methodology for assessing "depth of planning" will be presented and its use in these studies illustrated. Planning behavior in situations where there are no fixed procedures will be emphasized and associated data reported. Informal results for individual differences in planning style will be considered.

LANDING AIRPLANES, DETECTING TRAFFIC, AND THE DARK FOCUS

Stanley N. Roscoe
Behavioral Engineering Laboratory
New Mexico State University

ABSTRACT

Misjudgments of position in flight and failures to detect other airborne traffic are casualties of the eternal tug-of-war between visible texture and the pilot's dark focus. The eye is lazy and resists the pull of a distant stimulus, preferring to rest at a relatively short focal distance as it does in the dark or when looking at the sky. Results from 23 experiments show that judgments of apparent size are highly correlated with visual accommodation distance, and the difficulty of detecting airplanes on stationary collision courses is greatly aggravated when focus is trapped by structure close to the eyes. Subject, task, and environment variables all interact to determine what we think we see.

PROBLEM

In 1950 I discovered that pilots making landing approaches by periscope come in high and land long and hard, unless the image of the scene is magnified by about 25 percent (Roscoe, 1950; Roscoe, Hasler, & Dougherty, 1966). In 1973 Everett Palmer at NASA-Ames Research Center experimentally confirmed the common observation that pilots also make high approaches and long, hard landings in flight simulators with contact visual systems (Palmer & Cronn, 1973). Why is it that either real or virtual images projected at x1 magnification cause objects such as airport runways to appear smaller and farther away than when viewed directly?

In 1975 Robert Randle of NASA-Ames and I set out to find answers. We had the expert help of Robert Hennessy, now Study Director of the NRC Committee on Human Factors, and a gaggle of graduate research assistants at San Jose State in 1975-77, the University of Illinois in 1977-1979, and New Mexico State in 1979-81. Together we discovered a correlation of 0.9 or greater between the apparent, or perceived, size of objects subtending a given visual angle and an observer's visual accommodation--the distance to which the eyes are focused (Roscoe, 1979a, 1979b).

Perhaps it is not surprising that a relationship between perceived size and eye focus for distances well beyond the near limit of "optical infinity" went undiscovered for so long. Stimulus distances normally included in laboratory experiments do not approach those from which pilots view the world below them in flight. Indeed, few laboratory experiments in which eye focus was actually measured have involved distances of more than a few meters. But now that a strong relationship between the distance of eye focus and apparent size has been established, other mysteries of visual perception and illusion in flight become fair game for reinterpretation.

For example: Why do pilots making landing approaches over water at night toward a brightly lighted city consistently come in low (Kraft, 1978) and sometimes land in the bay short of the runway? This has happened in Tokyo, San Francisco, Los Angeles, Salt Lake City, and so it goes. Why is it that military pilots making ground attack runs so often fail to pull up in time and fly into the terrain in clear daylight? And why, in a group of pilots with "normal" vision, will some spot "bogies" so much sooner than others?

BACKGROUND

Before offering speculative answers to such questions, some background is in order. By 1970 Randle had developed a classical Pavlovian conditioning technique, employing automatic biofeedback of focusing responses, to study the extent of possible voluntary control of accommodation. Randle's initial purpose was to teach children how to avoid becoming myopic. Then during the early 1970s, Hennessy, working with his mentor, Herschel Leibowitz, and fellow graduate student, Fred Owens. at Penn State, greatly extended our understanding of the "anomalous" empty-field, night, and instrument myopias and clarified the role of the dark focus, or relaxed accommodation, of the eye (Hennessy, 1975; Leibowitz, Hennessy, & Owens, 1975; Leibowitz & Owens, 1975a, 1975b, 1978; Owens, 1976, 1979; Owens & Leibowitz, 1975, 1976a, 1976b).

Then, between 1975 and 1980 my students, colleagues, and I did 21 experiments involving the relationships among visual stimulus variables, eye accommodation, and associated perceptual responses (Roscoe, 1975; Roscoe, Olzak, & Randle, 1976; Roscoe, 1977; Iavecchia, Iavecchia, & Roscoe, 1978; Roscoe, & Benel, 1978; Benel & Benel, 1979; Simonelli & Roscoe, 1979; Benel, 1979; Gawron, 1979; Simonelli, 1979a, 1979b; Hull, Gill, & Roscoe, 1979; Roscoe, 1979a, 1979b, 1980a, 1980b; Gawron, 1980; Randle, Roscoe, & Petitt, 1980). Reports of two recent experiments are in preparation. There is no longer any question we have hit on a line of investigation of great importance to aviation.

Among the many findings, the following stand out as contributing to our understanding of why pilots often misjudge sizes and distances and fail to see and avoid other aircraft in flight:

1. Judgments of size, and by inference the distance, of objects in natural outdoor vistas are strongly dependent on the distance to which the eyes are focused ($r > 0.9$). The exact functional relationship is partially obscured by the unequal psychophysical units of the inverse dioptric scale, D = 1/focal distance in meters.

2. Accommodation to natural vistas depends in a complicated way on the dark focus of the individual, the retinal locus and spatial frequency of visible texture, and the sharpness of focus needed for the discrimination of object identity, for example, reading a sign.

3. Individual differences in dark focus range from perhaps 15 D (7 cm) in extremely myopic people to as distant as -4 D (far beyond "optical infinity") in the extremely hyperopic; the more distant the individual's dark focus, the greater his or her tendency to focus beyond an acuity target to maximize apparent size for the discrimination of detail (Simonelli, 1979a).

4. Some individuals can be trained more readily than others to control the focal distance of their eyes voluntarily; there is some evidence that such trainability depends in part on the individual's dark focus and that both the selection and training of pilots should take such characteristics into account.

THE MOON ILLUSION REVISITED

A convenient way to study perceptual responses to the distant vistas seen in contact flight is to use a technique developed by Lloyd Kaufman and Irvin Rock (1962) to quantify the moon illusion. By superposing a collimated disk of light on any natural outdoor or laboratory scene and providing an adjustable-diameter comparison disk nearby, surprisingly consistent estimates of the apparent size of the simulated "moon" can be obtained. Our adaptation of the Kaufman and Rock technique (known affectionately as "the moon machine") has been used in a series of experiments to correlate measured eye accommodation, judgments of apparent size, and characteristics of both natural and artificial visual scenes.

These experiments have shown that with both natural and artificial scenes, whether in daylight or at night, when viewing conditions cause the eyes to focus near, the moon shrinks, and when they cause distant focus, the moon grows (Iavecchia, Iavecchia, & Roscoe, 1978; Simonelli & Roscoe, 1979; Benel, 1979; Hull, Gill, & Roscoe, 1979). Whatever the causal explanation may turn out to be, this highly invariant relationship is the key to many of the misjudgments experienced by pilots. Such misjudgments can cause pilots to land in the water at night, fly into the terrain or overshoot a runway in the daylight, or fail to see and avoid another airplane on a collision course.

CRITICAL VARIABLES

Viewing "conditions" that cause shifts in focus either outward or inward from the dark focus include subject, task, and environment variables as well as the distribution of visible texture.

Subject Variables

Differences in perceptual abilities among people qualifying as having normal "20/20" vision are staggering (Simonelli, 1979a). Some are surprisingly near sighted while some have the ability to focus -4 D beyond "optical infinity," much like a zoom lens of a TV camera. An Air Force recruit, when told by Simonelli that he had remarkable vision, said, "Yes, Suh, I can tell the color of a frog's eyes at 100 paces." The recruit was not bragging; his acuity was on the order of 20/10 and his dark focus and far point well into the negative range.

Eye accommodation is a tug-of-war between the stimulus and the dark focus, with the stimulus normally pulling just hard enough to be seen and recognized. Nick Simonelli refers to this as the "acuity demand" of a stimulus. As we walk, drive a car, or fly low over the terrain, our accommodation is determined largely by Gibson's (1950) well-known "texture gradient." The retina unconsciously performs some kind of a product-moment averaging routine on the textural elements to reduce the blur, and the fact that much of the scene necessarily remains blurred normally goes unnoticed so long as the acuity demand remains low.

In daylight the gradient extends uninterrupted from the nose and other parts of the body to the near foreground and on out to the distant horizon. But from the cockpit at night, and even in daylight at higher altitudes, the gradient is not uninterrupted. Between nearby cockpit surfaces and the outside visible texture the gradient is interrupted by empty space. Even clouds are effectively textureless in that they present little acuity demand, and at night the outside texture is limited to a thin horizontal band of point light sources. Now this is where individual differences in dark focus can cause giant misperceptions.

If a pilot's dark focus is at about arms' length, normal for young healthy eyes, he will experience empty field myopia in daylight, as well as night myopia. Empty field myopia is reinforced by the stimulus pull of window posts and frames, some of which are even nearer than arms' length. In a recent experiment by Jan Hull at New Mexico State (Hull & Roscoe, in preparation), he found that while searching the sky, pilots focused at almost exactly the distance of interposed window posts, even when a post was no wider than the 2-1/2-inch distance between the eyes, thereby supposedly causing no binocular obstruction to outside vision.

Even though other traffic may be clearly visible, the effect of induced myopia is to blur the retinal image, reduce effective contrast, and make objects harder to see and apparently both smaller and farther away (Kraft, Farrell, & Boucek, 1970; Roscoe, 1979a, 1979b). Targets can still be detected, particularly if they move, flash, or glisten, or if they present an extended distinctive shape, such as a long, thin contrail. Unfortunately, another airplane on a collision course doesn't move, it only grows, and it must subtend a visual angle of more than 8 minutes if it is to be readily detected when badly out of focus.

In a second experiment soon to be reported, Hull and I tested the effects of window posts 2-1/2 and 4-5/8 inches wide, and 12 inches in front of the eyes, on the probability of detecting simulated contrails at various elevations projecting various angular distances from the right or left edge of a post. With a 2-1/2-inch post (the minimum visibility standard), the probability of detection in a single fixation of 1/3 second ranged from 0.79 to 0.97 as the angular length of a contrail increased from 6 to 16 degrees to the right or left of forward vision.

With the 4-5/8-inch post (2-1/8 inches greater than the interpupillary distance), the probabilities of detection for corresponding contrails plunged to 0.10 for 6 degrees (barely visible to one eye or the other at the right of left edge of the post) and gradually increased to 0.29 for 7 degrees, 0.55 for 9 degrees, 0.68 for 12 degrees, and 0.65 for 16 degrees. In addition to the total binocular obscuration caused by an oversized window post, the probability of detection of a contrail in the monocular penumbra on either side is greatly reduced. Few pilots are aware of the danger caused by wide window posts a few inches from the eyes.

Now for a different danger. If a pilot's dark focus is quite distant, possibly beyond optical infinity, and his attention is directed to the lights of a coastal airport and the city rising beyond, the visual scene can appear greatly magnified. The nearer lights of the runway threshold will expand downward from the horizontal band of city lights, thereby making it appear that the airplane is high on final approach. The pilot may compensate by reducing power and drop below proper glideslope. At some point the low position will suddenly become apparent, and normally the pilot will add sufficient thrust to land safely; but with engines spooled down, thrust may come too late to avert the water landing.

Task Variables

Pilots with normal, intermediate dark focus distances can experience similar outward shifts in accommodation when task demands create elevated workloads and stresses. Several experiments have demonstrated accommodation shifts with interposed mental activities (Malmstrom, 1978; Gawron, 1979; Randle, Roscoe, & Petitt, 1980). Bob Randle, John Petitt, and I, at Ames Research Center, found that accommodation shifted outward between measurements 20 seconds and 10 seconds before touchdown on simulated landing approaches by reference to a computer-animated night visual scene. We have no comfortable explanation.

However, outward accommodation is mediated by the sympathetic branch of the autonomic nervous system (Cogan, 1937; Benel, 1979; Gawron, 1979). That's the one that makes us run faster and fight harder. It also helps us see the distant enemy in the shadow of a rock and the deer behind a bush. It increases our acuity by magnifying what we see, just as outward focus magnifies the moon. Can it be that the flow of sympathetic adrenalin in the attack pilot expands his visual world, makes the ground appear lower, and causes him to pull up too late? When a periscope's magnification is set too high, pilots are often surprised by a touchdown far short of the runway (Roscoe, Hasler, & Dougherty, 1966).

Randle, Petitt, and I also found that pilots do not accommodate accurately to changing focus demands induced by ophthalmic lenses. They responded slightly better to a direct view of the computer-animated display than to collimated virtual images as presented on head-up displays, thus bringing into question the supposed advantage of preparing the eyes to see the runway when it suddenly appears on low visibility approaches. Optically collimating an image tends to release our focus from the distant stimulus and allows it to lapse toward the dark focus distance.

Environmental Variables

This discussion might have been headed, "St. Thomas Revisited." Surely one of the most puzzling and dramatic aviation mysteries surrounds the crash of an American Airlines B-727 at Harry S. Truman Field, St. Thomas, Virgin Islands, in 1976. Captain Arthur Bujnowski had made 154 uneventful landings on the same short, wide runway with similar daylight visibility and light, gusting winds. But on April 27, 1976, Art Bujnowski made a normal, "slotted" approach, leveled off a few feet above the runway, and floated beyond the point of no return. The flaming crash and resulting smoke cost the lives of 35 passengers and 2 flight attendants (NTSB, 1976; Roscoe, 1976, 1980a).

Art Bujnowski is a pilot's pilot, a skillful, calm, no-nonsense ex-Captain, now in forced retirement in Connecticut and permanently grounded. But three minutes before his ill-fated landing in 1976 he was in extreme pain from blocked ears due to an abnormal increase in cockpit and cabin pressure caused by mismanagement of the air compressor during a rapid descent. Other crew members and passengers were in similar pain. Intense stimulation of the inner ears causes an accommodative spasm of the eyes at about arms' length on average (Clark, Randle, & Stewart, 1975).

Neither Bujnowski nor his copilot could see the clearly visible VASI lights on final aproach, and both testified they expected the airplane to touch down, as doctrine called for, 1000 feet from the runway threshold. But it did not douch down, and with about 1500 feet of runway remaining, the copilot finally said, "You're still high, Art." I've made the same comment dozens of times as safety pilot in the right seat while a pilot-subject rounded out high and floated with the periscope set at x1 mangification. The visual field is compressed with near accommodation, and the runway appears higher than it is. As Art Bujnowski remembered:

> "...all I could see were cottages and stores or whatever they were. But it seemed like the activity was right there at eye level, ..." (Transcript of NTSB public hearing, p. 360.)

REAL-WORLD APPLICATION

Randle's (1970) demonstration of the possibility of conditioning the accommodation reflex by the application of biofeedback calls for systematic investigation of the trainability of individuals varying in dark focus distances and other oculomotor abilities. Basic data in this area are fragmentary but promising, and effective conditioning techniques are needed involving only simple, inexpensive equipment that can be used by instructors or technicians with limited training, or even by the individual pilot. We have had some success using a simple vernier optometer constructed from crosspolarized strips of cheap filter material (Simonelli, 1979b). This and other possible techniques need to be investigated.

The effective focal distance of the eyes can be manipulated either voluntarily, following bioconditioning, or involuntarily, by having pilots wear polyfocal glasses as is done by United Air Lines (Harper & Kildera, 1968). Acuity in resolving distant stimuli is enhanced by focusing at a distance greater than that of the stimulus to be discriminated. It is possible that detection of distant "point" targets, such as other aircraft, also can be enhanced by inducing accommodation to distances at or "beyond" optical infinity for individuals capable of unusually distant focus.

Each of the so-called anomalous myopias and its associated micropsia are encountered in varying degrees by pilots flying airplanes, particularly ones with head-up displays. Similar myopic responses and micropsic perceptions occur in airplane simulators with contact visual systems. Recall that it was our concern with the bias errors in landing with imaging flight displays that stimulated our interest in this line of research in the first place. We found that pilots do learn to compensate partially for such biased perceptions. The possibility of training individuals to recognize conditions in which to expect macropsic as well as micropsic misperceptions and to compensate for them voluntarily is out there like Mt. Everest.

REFERENCES

Benel, R.A. *Visual accommodation, the Mandelbaum effect, and apparent size* (Tech. Rep. BEL-79-1/AFOSR-79-5). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Benel, R.A. & Benel, D. C. R. *Accommodation in untextured stimulus fields* (Tech. Rep. Eng Psy-79-1/AFOSR-79-1). Urbana-Campaign, IL: University of Illinois, Department of Psychology, 1979.

Clark, B., Randle, R.J., & Stewart, J.D. Vestibular-ocular accommodation reflex in man. *Aviation, Space, and Environmental Medicine*, 1975, *46*, 1336-1339.

Cogan, D.G. Accommodation and the autonomic nervous system. *Archives of Ophthalmology*, 1937, *18*, 739-766.

Gawron, V.J. *Eye accommodation, personality, and autonomic balance* (Tech. Rep. BEL-79-2/AFOSR-79-6). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Gawron, V.J. *Differences among myopes, emmetropes, and hyperopes* (Tech. Rep. BEL-80-1/AFOSR-80-1). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1980.

Harper, C.R. & Kidera, G.J. Flight deck vision and the aging eye. *Aerospace Medicine*, 1968, *39*, 1119-1122.

Hennessy, R.T. Instrument myopia. *Journal of the Optical Society of America*, 1975, *65*(10), 1114-1120.

Hull, J.C., Gill, R.T., & Roscoe, S.N. *Locus of the stimulus to visual accommodation: Where in the world, or where in the eye?* (Tech. Rep. BEL-79-5/AFOSR-79-9). Las Cruces, New Mexico State University, Behavioral Engineering Laboratory, 1979.

Iavecchia, H.J., Iavecchia, H.P., & Roscoe, S.N. *The moon illusion: Apparent size and visual accommodation distance* (Tech. Rep. Eng Psy-78-4/AFOSR-78-3). Urbana, IL: University of Illinois at Urbana-Champaign, Engineering Psychology Laboratory, 1978.

Kaufman, L. & Rock, I. The moon illusion. *Scientific American*, 1962, *207*, 260-268.

Kraft, C.L. A psychophysical contribution to air safety: Simulator studies of visual illusions in night visual approaches. In H.A. Pick, Jr. (Ed.), *Psychology: From research to practice*. New York: Plenum Press, 1978.

Kraft, C.L., Farrell, R.J., & Boucek, G.P., Jr. *Visual performance from the 737 cab as a function of pilot's position within the cab*, (Unnumbered internal report). Seattle: The Boeing Company, 1970.

Leibowitz, H.W. & Owens, D.A. Anomalous myopias and the intermediate dark focus of accommodation. *Science*, 1975, *189*, 646-648. (a)

Leibowitz, H.W. & Owens, D.A. Night myopia and the intermediate dark focus of accommodation. *Journal of the Optical Society of America*, 1975, *65*, 1121-1128. (b)

Leibowitz, H.W., Hennessy, R.T., & Owens, D.A. The intermediate resting position of accommodation and some implications for space perception. *Psychologia*, 1975, *18*(3), 162-170.

Malmstrom, F.V. Effects of concurrent mental activity on static and dynamic accommodation responses (Doctoral dissertation, Oklahoma State University, 1978). Dissertation Abstracts International, 1979, 39B, 61-62.

National Transportation Safety Board (NTSB). Aircraft accident report: American Airlines, Inc., Boeing, 727-95, N1963; St. Thomas, Virgin Islands, April 27, 1976. Washington, DC: United States Government Printing Office, Report NTSB-AAR-77-1, 1976.

Owens, D.A. Factors influencing steady-state accommodation (Doctoral dissertation, Pennsylvania State University, 1976). Dissertation Abstracts International, 1977, 37B, 5863.

Owens, D.A. The Mandelbaum effect: Evidence for an accommodative bias toward intermediate viewing distances. Journal of the Optical Society of America, 1979, 69, 646-652.

Owens, D.A. & Leibowitz, H.W. The fixation point as a stimulus for accommodation. Vision Research, 1975, 15, 1161-1163.

Owens, D.A. & Leibowitz, H.W. Oculomotor adjustments in darkness and the specific distance tendency. Perception and Psychophysics, 1976, 20(1), 2-9. (a)

Owens, D.A. & Leibowitz, H.W. Night myopia: Cause and a possible basis for amelioration. American Journal of Optometry and Physiological Optics, 1976, 53, 709-717. (b)

Palmer, E. & Cronn, F.W. Touchdown performance with a computer graphics night visual attachment. Proceedings of the AIAA Visual and Motion Simulation Conference. New York: American Institute of Aeronautics and Astronautics, 1973.

Randle, R.J. Volitional control of visual accommodation. AGARD Proceedings No. 82 on Adaption and Acclimatization in Aerospace Medicine. Neuilly-sur-Seine, France: North Atlantic Treaty Organization, 1970.

Randle, R.J., Roscoe, S.N., & Pettit, J. Effects of accommodation and magnification on aimpoint estimation in a simulated landing task (Technical Paper NASA-TP-1635). Washington, DC: National Aeronautics and Space Administration, 1980.

Roscoe, S.N. Aircraft pilot performance as a function of the extent and magnification of the visible horizon (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1950). Dissertation Abstracts International, 1951, 11, 173.

Roscoe, S.N. Ground-referenced visual orientation in flight control tasks: Judgments of size and distance. In Vision R&D Program Review. Washington, DC: Naval Air Systems Command and Office of Naval Research, 1975, 145-174. [Originally issued by the University of Illinois as Technical Report ARL-75-7/ONR-75-1, 1975.]

Roscoe, S.N. Appendix I: Human factors and crew performance in the St. Thomas accident. In Accident investigation post-hearing submission to the National Transportation Safety Board, B-727 accident, St. Thomas, Virgin Islands, April 27, 1976. Arlington, TX: Allied Pilots Association, 1976.

Roscoe, S.N. How big the moon, how fat the eye? (Tech. Rep. ARL-77-9/AFOSR-77-8). Savoy, IL: University of Illinois at Urbana-Champaign, Aviation Research Laboratory, 1977.

Roscoe, S.N. When day is done and shadows fall, we miss the airport most of all. *Human Factors*, 1979, *21*, 721-731. (a)

Roscoe, S.N. *Ground-referenced visual orientation with imaging displays: Final report* (Tech. Rep. Eng Psy-79-4/AFOSR-79-4). Champaign, IL: University of Illinois at Urbana-Champaign, Department of Psychology, 1979. (b)

Roscoe, S.N. *Aviation psychology*. Ames, IA: Iowa State University Press, 1980. (a)

Roscoe, S.N. Bigness is in the eye of the beholder. *Proceedings of the Seventh Symposium on Psychology in the Department of Defense*. Colorado Springs: United States Air Force Academy, 1980, 39-80. (b)

Roscoe, S.N. & Benel, D.A. *Is the eye smart or the brain forgiving?* (Tech. Rep. Eng Psy-78-1/AFOSR-78-1). Champaign, IL: University of Illinois at Urbana-Champaign, Department of Psychology, 1978.

Roscoe, S.N., Hasler, S.G., & Dougherty, D.J. Flight by periscope: Making takeoffs and landings; the influence of image magnification, practice, and various conditions of flight. *Human Factors*, 1966, *8*, 13-40. [Originally issued in 1952 by the Office of Naval Research as Human Engineering Report 71-16-9, classified CONFIDENTIAL.]

Roscoe, S.N., Olzak, L., & Randle, R.J. Ground-referenced visual orientation with imaging displays: Monocular versus binocular accommodation and judgments of relative size. *Proceedings of the AGARD Aerospace Medical Specialists Meeting, Athens, Greece*. Neuilly-sur-Seine, France: North Atlantic Treaty Organiziation, 1976.

Simonelli, N.M. *The dark focus of visual accommodation: Its existence, its measurement, its effects* (Tech. Rep. BEL-79-3/AFOSR-79-7). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979. (a)

Simonelli, N.M. *Polarized vernier optometer* (Tech. Rep. BEL-79-4/AFOSR-79-8). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979. (b)

Simonelli, N.M. & Roscoe, S.N. *Apparent size and visual accommodation under day and night conditions* (Tech. Rep. Eng Psy-79-3/AFOSR-79-3). Champaign, IL: University of Illincis, Department of Psychology, 1979.

THE DARK FOCUS OF ACCOMMODATION AND PILOT PERFORMANCE

Russell A. Benel and Thomas L. Amerson, Jr.

Essex Corporation
Alexandria, Virginia

ABSTRACT

Two closely related lines of research are presented in this paper. The first follows directly from the work on the resting state of visual accommodation (dark focus) as a predictor of visual target acquisition and perceptual judgments. The second describes systematic (diurnal) variations in the dark focus. The possible relationship between this research and available performance (accident) data is described. A discussion of the impact of this human capability factor upon function allocation in aviation systems is included with particular attention to selection criteria, training, and ameliorative (hardware) techniques. In addition, the relationship between the dark focus (as a preferred focal distance) and system monitoring is considered both for aircraft and for ground-based systems.

INTRODUCTION

Flight is heavily dependent upon visual acquisition of information. In commercial, military, and civil aviation, apparently stringent standards for vision are applied. Despite a long history of the application of visual standards, errors in acquisition of information remain relatively common. One such error is failure to acquire necessary information under unobscured circumstances. The most prevalent situation is air-to-air search where targets remain unseen until quite close. This decreases the margin for avoidance and, in the military situation, decreases the possibility for a first strike. Misperception of visual information is also common in aviation. The work of Roscoe and associates in the early 1950's with imaging displays provided evidence for systematic biases in landing performance. More recent evidence, again from Roscoe and associates, indicates similar effects occur during direct viewing of the outside world. These errors in visual acquisition are not trivial, having been implicated in the loss of lives and in millions of dollars in property damage.

The basic visual capabilities of men need to be fully considered. Any such consideration should depend upon empirical evidence, rather than conventional wisdom and/or historical precedent. Periodically, the "facts" as we know them need to be reexamined in light of current evidence. Frequently, the evidence supports the status quo, but many times continued support is unwarranted. This has been the case for one basic attribute of the human visual system. The assumed resting position for focus or accommodation of the eye is not at the far point, as has been assumed since Helmholtz; rather, the weight of evidence indicates that a dark focus (labelled such for the most convenient circumstances in which to measure unstimulated accommodation) is on average within arm's length, with an appreciable variability in this distance among individuals.

EXPERIMENT I

A body of research has been accumulating which strongly supports an intermediate dark focus and indicates concomitantly a variety of "normal" viewing situations which are influenced by this tendency. One effect which has been documented involves vision through interposed surfaces. Whiteside (1957) observed that some pilots focused spots on the aircraft canopy during air-to-air search, but Mandelbaum (1960) was apparently the first to document the circumstances under which the interposed texture, rather than the desired object, was focused. Owens (1979) found that the "Mandelbaum effect" is greatest when the interposed screen and the observer's dark focus correspond, i.e., the eye exhibits an inherent bias toward objects presented near the dark focus.

In work alluded to earlier, Roscoe and associates have found a similar intermediate tendency in visual accommodation when subjects view natural scenes of varying texture or "plain" versus "fancy" targets (for a short review see Roscoe, 1981). The accuracy of accommodation can be shown to vary with a great many factors including: (1) distance to a surrounding texture (Hennessy & Leibowitz, 1971); (2) background appearance (Benel & Benel, in press); stimulus characteristics (Simonelli, 1979); and even the mood of the observer (Miller, 1978).

An additional difficulty arises with inappropriate accommodation. If one views monocularly a relatively distant object and then focuses on one finger held in front of the active eye, the object will shrink in apparent size. In fact, moving the finger will make the object appear to change size, shrinking when movement is toward the eye and expanding when away (attributed to Hoffman; see Ogle, 1950). Thus, shifts in accommodation may result in more than a loss of acuity. Shifts in apparent size (and apparent distance) may interfere with safe operation of vehicles. This is particularly so in aircraft where the requirement for accurate perception both for traffic avoidance (optimal acuity) and landing (optimal size/distance judgment) is extremely high and the penalty for failure to maintain these standards would be severe.

Benel (1979) conducted a series of experiments culminating in a demonstration of the effects of interposed texture on size perception of the image of a distant object. Other findings from this series included a replication of earlier work on the Mandelbaum effect. Namely, the nearer to the dark focus distance that an interposed surface lies, the more likely it is that accommodation will be "trapped" at that distance. Also, the more heavily textured the surface, the more likely it is that it will exert this effect. A third experiment verified Hoffman's demonstration of shifts in size, but only qualitatively, i.e., only larger or smaller. The last experiment to be described below quantified to some extent these latter findings.

Method

Subjects. Twelve observers were selected with a minimum uncorrected near and far acuity of 20/25.

Apparatus. Acuity measurements were made with an Orthorater. Accommodation measurements were made with a laser optometer (for a complete description and operating procedures, see Hennessy & Leibowitz, 1972). Briefly, laser light reflected from a diffusing surface is granular or speckly in appearance. If the surface is a rotating drum, the apparent motion of the speckles allows accommodation to be measured. If the motion corresponds to the direction of rotation, the eye is overaccommodated (focussed nearer). If against the direction, it is underaccommodated. If the drum is optically conjugate to the retina, no consistent movement is seen. The Badal principle allows the effective range of the optometer to be accomplished in a short space. Placing a positive lens in the reflected path one focal length from the lens of the eye allows the optical distance to be varied from optical infinity to the effective power of the lens with no changes in size or brightness of the image (Ogle, 1971). Exposures were limited to .5 sec, a length of time shown to have no differential effect on measured accommodations. A schematic of this arrangement may be seen in Figure 1.

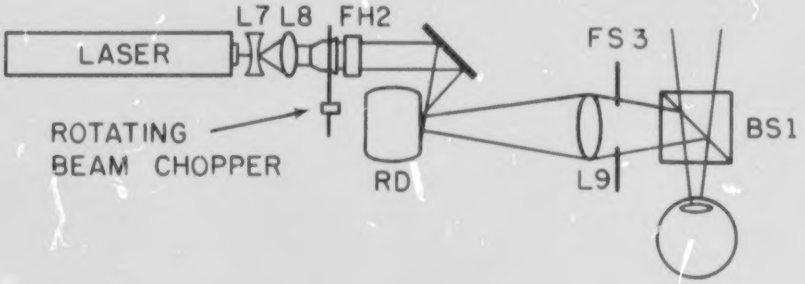


Figure 1. Schematic diagram of the laser optometer

The size-matching apparatus (described in detail by Iavecchia, Iavecchia, & Roscoe, 1978) provides alternate viewing of a collimated, transilluminated disc superimposed on a selected field of view and a variable diameter transilluminated disc projected from one meter with a black background. The former served as a standard to be matched by the latter. A 37° field of view was available. A chin and forehead rest maintained a constant eye position. Cut-away schematics of the apparatus are shown in Figure 2. The laser was mounted to the left of the apparatus and its speckle pattern was observed in a beamsplitter. Black fiberglass mesh window screening served as the interposed texture.

Procedure. Observers were instructed to make the adjustable moon appear the same size as the collimated moon. For each measurement the observer was allowed to refer to the collimated moon a maximum of three times.

Preceding and following each session, the dark focus was measured for each observer. The session began and ended with two accommodation and apparent size matches to the external scene and moon without the screen. The screen was presented twice sequentially at each of four distances (.75, 1.5, 2.25, and 3.0 D) and accommodation and apparent size measures were collected at each distance. The order of presentation was counterbalanced across observers. The screen was moved while the mirror reflecting the adjustable disc occluded the external scene. This prevented the observer from knowing with any degree of certainty the precise distance to the screen.

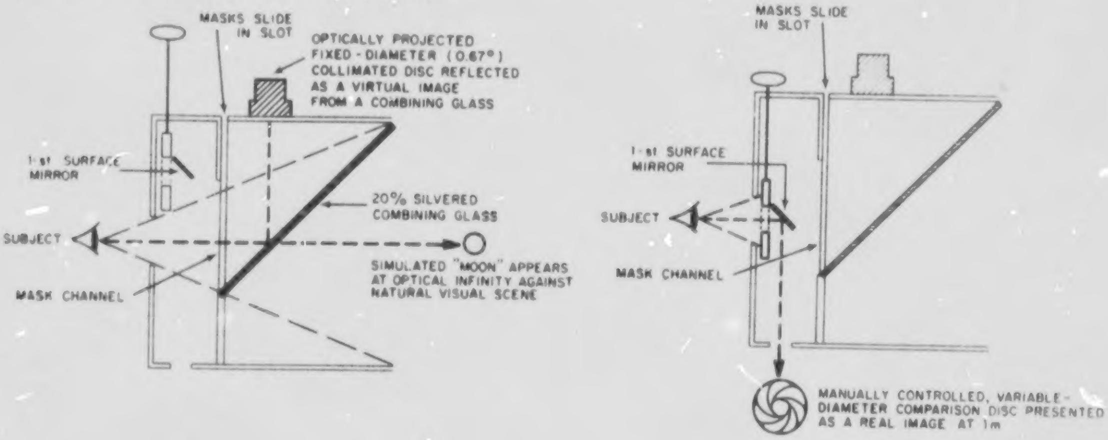

MASKS SLIDE IN SLOT
OPTICALLY PROJECTED FIXED-DIAMETER (0.67°) COLLIMATED DISC REFLECTED AS A VIRTUAL IMAGE FROM A COMBINING GLASS
1-st SURFACE MIRROR
20% SILVERED COMBINING GLASS
SUBJECT
SIMULATED "MOON" APPEARS AT OPTICAL INFINITY AGAINST NATURAL VISUAL SCENE
MASK CHANNEL
MASKS SLIDE IN SLOT
1-st SURFACE MIRROR
SUBJECT
MASK CHANNEL
MANUALLY CONTROLLED, VARIABLE-DIAMETER COMPARISON DISC PRESENTED AS A REAL IMAGE AT 1m

Figure 2. Cut-away schematics of the stimulus presentation box. On the left, box is set for viewing scene with superposed collimated lighted disc, on right for viewing adjustable comparison disc.

Results

Two 2-way ANOVAs summarized in Tables 1 and 2 were performed on the accommodation and apparent size data. Screen presentation distance had a reliable effect on both accommodation and apparent size. The relationship between accommodation and apparent size was analyzed by a series of Pearson product-moment correlations. For the initial control situation (no screen), accommodation and apparent size correlated reliably ($\underline{r} = -.56$, $p < .05$), indicating apparent size decreased with nearer accommodation. The same relationship was found across screen conditions ($\underline{r} = -.76$, $p < .01$).

TABLE 1

ANALYSIS OF VARIANCE FOR ACCOMMODATION

SOURCE	SS	DF	MS	F
Screen Distance (A)	4.459	5	.892	8.13***
A x B	6.026	55	.110	
Observers (B)	27.664	11	2.515	

TABLE 2

ANALYSIS OF VARIANCE FOR APPARENT SIZE

SOURCE	SS	DF	MS	F
Screen Distance (A)	10870.06	5	2174.01	6.35***
A x B	18819.76	55	342.18	
Observers (B)	72926.81	11	6629.71	

*** $p < .001$

When data are pooled across observers, the correlation between apparent size and accommodation is -.96 (see Figure 3). Converted to distance of accommodation (in meters) rather than diopters, the correlation appears more linear and is positive at .98.

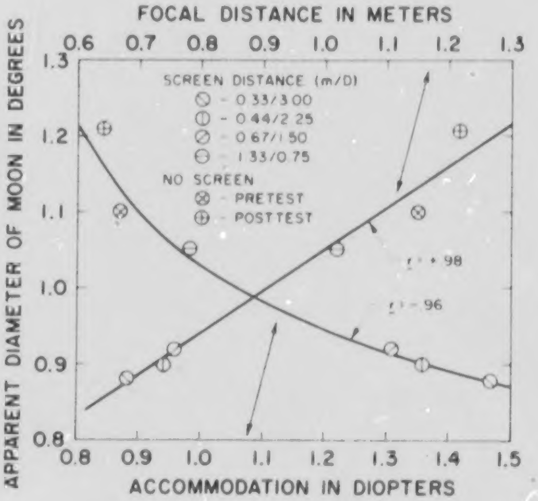


Figure 3. Correlation between apparent size and accommodation.

Discussion

The monotonic increase in mean accommodation and decrease in mean apparent size as the screen position approached the observer combined with the relatively high correlation between apparent size and accommodation suggest that the distance to which an observer is accommodated is a real, quantifiable correlate of apparent size. Although accommodation is not to be taken as an exclusive factor, the magnitude of the correlations between apparent size and accommodation found indicates that a large portion (approximately 50% in the current study) of the total variance may be predicted by accommodation. Furthermore, the ratio of size judgments under standard conditions (no screen) to those with the near screen indicates a magnification factor of 1.25 is needed to equalize image sizes. This figure is in agreement with the literature on imaging displays for landing aircraft.

EXPERIMENT II

It has been suggested that oculomotor function, like other body functions, varies by time of day. In a ten-day study by (Leibowitz, 1976), the dark focus of three male subjects displayed a possible diurnal variation. The "nonstressed" subject typically showed dark focus measures .2 - .3 D less than morning measures of the same day. The "stressed" subjects exhibited a reversed relationship. Thus, a diurnal effect may have been present as well as a stress effect. Miller (1978) has also reported a possible diurnal variation among subjects whose dark focus was measured morning and evening, three days per week, for three weeks. All 21 subjects have larger dark focus values in the evening. Subjects revealed a positive correlation among changeability of mood, variability of measured dark focus, and the amount of increase of accommodation in the evening dark focus measurement. Although the combination of these two works suggests a diurnal variation in the dark focus, the specific magnitude and nature of this variation was indeterminable.

A study by Mershon and Amerson (1980) has shown the change in dark focus to be less than ± .3 D for subjects measured twice over a one-week interval and controlled for time of day. Thus, it can be assumed that the dark focus measures of a population, however variable by time of day, are relatively stable for a given time of day.

Recently, Amerson and Mershon (Amerson, 1980) collected data specifically to determine if (and to what extent) time of day variation occurs in the dark focus of accommodation.

Method

Subjects. Thirty-six male college students were screened for near and far visual acuity of at least 20/30 in each eye when wearing their usual correction.

Apparatus Acuity measurements were made with a Keystone Telebinocular. The dark focus of accommodation was measured by a laser optometer, similar to that previously described. This instrument allowed measurement of accommodative distances from 0.33 to 3.6 D (from 2.33 to 5.6 D with the addition of a -2 D lens in the subject's spectacle plane).

Procedure. The data collection followed a 6 x 6 design where each of the six first measurement sessions in the morning was paired with each of the six afternoon and evening sessions. Each subject came to the laboratory twice: in the morning between 8:30 and 11:20 a.m., and then later in the day at 1:00, 2:55, 4:50, 6:45, 8:40, or 10:35 p.m. A test included two measurements on the right eye.

Results

Time-of-day variations were tested by comparing the change scores for the morning measures and the return measures (taken between 1:00 and 10:30 p.m.).

The mean dark focus value for the first test was 2.02 D, similar to that found in a previous study with the same equipment (Mershon & Amerson, 1980). It should be noted that mean dark focus values may vary greatly even with large sample sizes of equivalent acuity (Simonelli, 1979).

Since the subjects came to the laboratory at one of six times during the morning, the baseline measures in the study were, themselves, subject to time-of-day variation. To test for homogeneity of the six groups for baseline measures of dark focus (DF 1), the Kruskal-Wallis One-Way Analysis of Variance was performed. The lack of differences among the six groups ($p > .50$) was striking.

For the dark focus, there was a significant variation in the shifts from the morning baseline measurements. In Figure 4, each point represents one subject's change in dark focus. Inward shifts are indicated by values below the dashed baseline. The sloping line connects the median shift values across the six return times. Thus, for some groups, the second dark focus measures were significantly different from the baseline measures. The pattern of these variations is consistent with expectations based on autonomic function. That is, there was an outward shift in dark focus value from morning until some time in the early afternoon, reflecting heightened sympathetic activity. An inward shift occurred in the dark focus during the afternoon and extended into the evening, reflecting heightened parasympathetic activity.

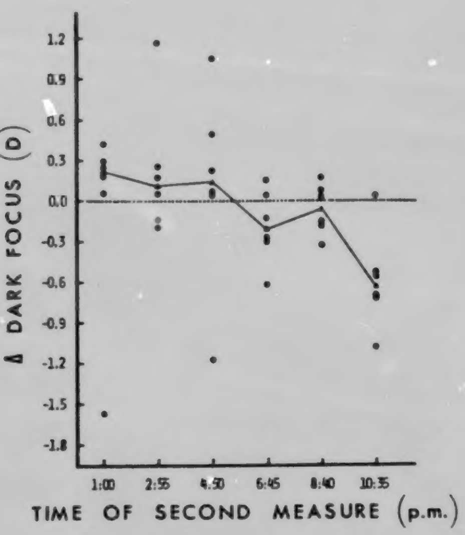


Figure 4. Individual shift in dark focus (DFI - DF2) by time of day (n = 36). Median group shifts are given by the connected triangles.

Discussion

Significant time-of-day shifts in the dark focus of accommodation were found. These shifts followed the pattern expected in that the power of the lens decreased from morning to early afternoon, whereupon a change in the amount and direction of shift commenced. By nighttime, the dark focus was observed to be greater (the lens more powerful) than in the morning.

GENERAL DISCUSSION

It is generally recognized that approaches and landings are among the most demanding phases of flight and are probably most associated with accidents and incidents. According to Hartman and Cantrell (1968) well over half of all accidents occur during the night. Of course, weather is implicated in large proportion of accidents but it is not differentially important across times of the day. Many factors might account for night/weather accidents, e.g., reduced visibility, increased pilot fatigue, disrupted circadian rhythms, etc. The relationship between such factors and the findings of the current research can be made explicit.

The research described in Experiment I indicates that accommodation to very adequate visual scenes may be disrupted by interposed texture (e.g., windscreens). Under situations with reduced external cues, this interposed texture becomes a more imperative stimulus. Thus, weather which reduces visibility may increase the tendency to focus at nearer texture, thereby reducing the likelihood that external cues will be perceived adequately. Likewise, the inaccurate focal state in evidence during the (predominantly) empty field of high altitude flight reduces the probability of detection of airborne targets (either traffic or bogies). One need not have too many San Diego incidents to see the relevance of optimal traffic detection.

Taking into account the results of Experiment II, it is obvious that the tendency for the dark focus distance to approximate the distance to a windscreen will vary diurnally. For an average pilot, who would be likely to have a distant dark focus (cf. Simonelli, 1979), the tendency to have a nearer dark focus should be highest during the period 0200-0600. (However, this expectation was not specifically tested.) This correlation does not imply causation, but it is highly suggestive of a contributing factor. If in fact these diurnal variations in the dark focus result in differential perceptual events, then the subjective appearance of objects such as runways will differ as well. The potential to respond differently and inappropriately is certainly real.

As systems become automated (albeit with manual override) the predominant role for man shifts from manipulator to monitor. The perceptual abilities of man will be taxed heavily. It appears likely that perception of approaches in automated landing systems will be subject to systematic errors (Randle, Roscoe, & Pettit, 1980). These errors may be particularly prevalent when approaches are made under circumstances which depart significantly from the normal time of day and/or visibility conditions. An additional factor in system monitoring performance relates to the dark focus as a preferred focal distance and the optimal distance to visual displays. Apparently, the least strain and best performance should result when the observer/display distance closely approximates the dark focus.

Since the dark focus is an individual characteristic with reasonable stability, it would be possible to develop a selection criterion based on the dark focus for various operator positions. For example, fighter pilots who would profit immensely from earlier air-to-air target detection could be selected on the basis of a distant dark focus. Conversely, when selection is impossible or undesirable, corrective lens could be prescribed to "correct" the dark focus either beyond the windscreen or to correspond to the observer/display distance. In systems needing both near and far optimization, operators would require multifocal lenses. An additional possibility is training of voluntary accommodation, a technique which has shown modest promise.

System design is performed currently with little regard to the dark focus of the eventual user population. Fortunately, many rules of thumb for display distance, etc., do lead to designs with observer/display distances that approximate the average dark focus. Unfortunately, many windscreen distances are also at that distance. Windscreens themselves have come under increasing scrutiny. It is likely that additional research will generate an increased emphasis on design in this area.

Although much evidence currently exists, many critical questions remain. For example, does the dark focus maintain its robustness as a determining factor in accommodation across the day? Does this variation follow some orderly circadian pattern? Would selection and/or training improve first strike probability in fighter engagements? Can corrective lenses reduce bias in landing judgments? Can corrective lenses and/or equipment placement reduce complaints about visual display induced fatigue? Further research should clarify these issues and provide definitive guidance for selection, training, and design.

REFERENCES

Amerson, T. L., Jr. Time of day variations in the dark focus of accommodation and dark convergence. Unpublished MS Thesis, North Carolina State University, 1980.

Benel, R. A. Visual accommodation, the Mandelbaum effect, and apparent size (Tech. Rep. BEL-79-1/AFOSR-79-5). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Benel, R. A. & Benel, D. C. R. Background influences on visual accommodation: Implications for target acquisition. Paper sumbmitted for publication.

Hartman, B. O. & Cantrell, G. K. Psychologic factors in "landing-short" accidents. Flight Safety, 1968, 2, 26-32.

Hennessy, R. T. & Leibowitz, H. W. The effect of peripheral stimulus on accommodation. Perception and Psychophysics, 1971, 10, 129-132.

Hennessy, R. T. & Leibowitz, H. W. Laser optometer incorporating the Badal principle. Behavioral Research Methods and Instrumentation, 1972, 4, 237-239.

Iavecchia, H. J., Iavecchia, H. P., & Roscoe, S.N. The moon illusion: Apparent size and visual accommodation distance (Tech. Rep. Eng Psy-78-4/AFOSR-78-3). Urbana, Illinois: University of Illinois at Urbana-Champaign, Engineering Psychology Laboratory, 1978.

Leibowitz, H. W. Visual perception and stress. In C. Borg (Ed.) Physical work and effort. New York: Peranmon, 1976.

Mandelbaum, J. An accommodation phenomenon. Archives of Opthalmology, 1960, 63, 923-926.

Mershon, D. & Amerson, T. L., Jr. Stability of measures of the dark focus of accommodation, Investigative Opthalmology and Visual Science, 1980, 19, 217-221.

Miller, R. B. Mood changes and the dark focus of accommodation. Perception and Psychophysics, 1978, 24, 437-443.

Ogle, K. N. Binocular Vision. Philadelphia: Saunders, 1950.

Ogle, K. N. Optics: An introduction for ophthalmologists. Springfield, Illinois: Thomas, 1971.

Owens, D. A. Factors influencing steady-state accommodation (Doctoral dissertation, Pennsylvania State University, 1976). Dissertation Abstracts International, 1977, 37B, 5863.

Owens, D. A. The Mandelbaum effect: Evidence for an accommodative bias toward intermediate viewing distances. Journal of the Optical Society of America, 1979, 69, 646-652.

Randle, R. J., Roscoe, S. N., & Pettit, J. Effects of accommodation and magnification on aimpoint estimation in a simulated landing task (Technical Paper NASA-TP-1635). Washington: NASA, 1980.

Roscoe, S. N. Landing airplanes, detecting traffic, and the dark focus. Paper presented at the 1981 Symposium on Aviation Psychology, The Ohio State University, April 21-22, 1981.

Simonelli, N. M. The dark focus of visual accommodation: Its existence, its measurement, its effects (Tech. Rep. BEL-79-3/AFOSR-79-7). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Whiteside, T. C. D. The problems of vision in flight at high altitude (AGARDograph No. 13). London: Butterworths Scientific Publications, 1957.

FUNCTIONAL OPTICAL INVARIANTS:

A NEW METHODOLOGY FOR AVIATION RESEARCH

Rik Warren and Dean H. Owen

The Ohio State University

ABSTRACT

The application of Gibson's (1979) "ecological approach to visual perception" to aviation psychology entails the use of information rich visual displays that must adequately and unambiguously enable a pilot to perform flight maneuvers. Optical information often takes the form of invariant properties of a changing optic array and functional invariants are defined as psychologically effective optical invariants. Their effectiveness is determined by empirical test but standard experimental paradigms are shown to be inappropriate for testing the effectiveness of information in rich displays due to the presence of inherent and unavoidable confounding factors that are here termed "secondary independent variables" in contradistinction to the "primary independent variables" manipulated by the experimenter. Recommendations for a new methodology and statistical treatment are offered and the implications for aviation psychology are discussed.

INTRODUCTION

The concept of functional optical invariants and the new methodology they entail were developed to meet certain difficulties we encountered in our attempt to apply J. J. Gibson's (1979) "ecological approach to visual perception" to fundamental problems of aviation psychology. Specifically, we are attempting to determine and describe the necessary and sufficient optical conditions that induce a perception of egomotion (self-motion). A knowledge of the necessary and sufficient optical bases for the perception of egomotion is needed to optimally design visual flight simulators and simulator training programs. Optimization is psychologically and economically important since underdesign results in poorer simulation training than possible and overdesign results in overly expensive training.

Ecological Optics and Optical Invariants

Since the concept of functional optical invariants is an extension of Gibson's (1979) theory, his ecological approach will be briefly reviewed. "Ecological optics" is the study of the information available in light and its origins trace back to Gibson's (1947) research on pilot selection and training in World War II. The principles of ecological optics that are relevant here are:

1. The light coming to a moving point of observation is structured owing to the structure of the environment and the observer's travel.

2. The optical structure is constantly changing, again owing to the observer's travel and also to events in the environment.

3. Over the changing structure or transformations of optical structure, there remain properties (often higher-order relationships) that do not change and are thus invariant over the transformation.

4. These optical invariants are claimed to be, or to form the bases of, the univocal information used by active perceivers to survive in and to exploit their environment.

Examples of change of optical structure. A common type of change of optical structure is the total change in the optical location or direction of points in the environment that corresponds to a displacement of the point of observation (Gibson, Olum, & Rosenblatt, 1955). Another example of change of optical structure is the change in optical size and optical density of environmental features due to a change in altitude.

Examples of optical invariants. During rectilinear egomotion. the optical position of the horizon is invariant over the otherwise total flow transformation. Also, the optical position of the ground point toward which a plane is heading is invariant if the path slope is constant. Since path slope (if there is no wind) is the ratio of the descent rate to the forward velocity, this means that the optical position of the aim point is furthur invariant over changes of descent and forward velocities as long as these change proportionately. Changes in these velocities do result in a change in the global optical flow rate (Warren, Note 1).

This example of path slope as a ratio of two rates of change underscores a common finding of ecological optics: often optical invariants emerge as rates of change during changes and especially as ratios of rates of change of environmental variables.

It is important to note that whether or not an optical invariant is indeed mathematically capable of specifying its source is a question for geometry; whether or not a particular optical invariant is actually used by an observer is a question for psychology. Hence, ecological optics is not itself a theory of perception, but a propaedeutic for one.

Perception and Functional Optical Invariants

Perception is defined as the pickup of information available in light. However, the existence of potentially available information does not force perceiving since, for example, an observer may not be attending or not yet have developed sufficient pickup skills (E. J. Gibson, 1969). Thus, optical invariants fall into two functional equivalence classes: those that are not utilized and are thus perceptually ineffective, and those that are indeed picked up and are thus perceptually effective.

Definition: Functional optical invariant. A functional optical invariant is an optical invariant that is perceptually effective (Owen, Warren, Jensen, Mangold, & Hettinger, in press). The term "functional" carries two implications: that of being used or utilized and also that of utility or practical, survival value.

The implication of being used means that the ultimate determination of whether or not something is a functional optical invariant is by empirical testing. This in turn implies that an adequate research methodology must be available.

The implication of utility means that the problems selected for study are motivated by practical concerns. This in turn implies that the research methodology be sensitive to the requirements of ecological validity.

Ecological functionalism and direct perception. The emphasis on ecological validity is a hallmark of the ecological approach and there are currently two active branches of development: One branch emphasizes the epistemological implications of the ecological approach and is associated with the term "direct perception" (e.g., Shaw & Bransford, 1977); another branch emphasizes the empirical implications and is termed "ecological functionalism" (Owen et al., in press).

This paper is on ecological functionalism and is concerned with the problem of how to study sensitivity to optical invariants. If standard experimental paradigms were adequate for testing candidates for functional optical invariant status, then this paper would be unnecessary. Unfortunately, standard experimental paradigms used today make assumptions that are inappropriate for perceptual research in aviation.

Assumptions of Standard Experimental Paradigms

The standard experimental paradigms we are refering to attempt to assess the effects on a performance dependent variable of systematic manipulation of two or more independent variables (IV) in a balanced, orthogonal factorial design. In practice, several assumptions are made in applying these paradigms to research problems. One class of assumptions may be termed "technical" and is not of interest here. These include the assumptions of random assignment and homoscadacticity. The second class of assumptions is concerned with the adequacy of the selection and evaluation of the IVs and are necessarily problem or context sensitive. In discussing these assumptions, the specific context is that of perceptual factors in aviation. The assumptions commonly made in current research are:

Assumption 1. It is assumed that the IVs generally selected are indeed the most relevant or germane for perception and action. Most relevant is used synonymously with directly relevant in a causal chain sense. For example, a common variable in the study of the perception of egospeed is actual speed of travel. The selection criterion apparently used is that of face validity albeit intuitively or tacitly applied.

Assumption 2. In any experiment, the total variation in the dependent variable may be partitioned into that due to: (a) the effects of the IVs selected and their interactions, (b) other systematic effects of either identified or unidentified sources, (c) individual differences, and (d) random error. Often, the sources of systematic effects may be intercorrelated so that advanced techniques such as multiple regression and correlation are required to evaluate the contribution of redundant factors, and hence interpretation is difficult (Cohen & Cohen, 1975).

But, it is assumed that the variation due to "other systematic effects" may be reduced to zero by means of a well designed and executed balanced orthogonal design. By well designed and executed is meant that the effects of all non-experimental factors (either identified or not) are made irrelevant by such means as elimination, use of a single level if elimination is not feasable, randomization, or counterbalancing so that their effects are self-cancelling and/or equal to zero. In essence, a well designed experiment is assumed to control for or be free of confounding factors. Technically, a confounding factor is a non- experimental or non-manipulated factor which has a non-zero coefficient of multiple determination (R^2) or curvilinear determination (η^2) with some IV or interaction of IVs of interest. It is further assumed that the presence of a confounding factor indicates a poor experiment.

Assumption 3. The third assumption is that data analysis is complete once an analysis of variance or regression analysis has rendered a verdict on the main effects and interactions. (Post hoc tests, trend analyses, and regression equations are included in the above analyses.) The main point here is that although the discovery of an interaction may lead to joy if it was predicted, or anguish if it was unexpected and "must be explained", it is assumed that no further explication as to just exactly how the variables combine is required. An interaction is defined as an effect beyond the mere addition of the effects of main factors, and there is no presumption that the exact mathematical nature of the non-additivity must be explicated. More serious is the assumption that main effects are terminal findings especially if no significant interaction is found.

Ecological Critique of Standard Methodology

As reasonable as the above assumptions are, they are not immune to criticism. One obvious critique of most experiments from the ecological viewpoint is the lack of ecological validity of the tasks and situations commonly used. But ecological validity does not concern us here since it is orthogonal to the procedural assumptions at issue.

Critique of Assumption 1

Perception exists for the purpose of acting in and on the environment. Hence it is reasonable to vary environmental conditions to determine their effect on perception and performance. But perception as the pickup of environmental information contained in light is perforce constrained by the available information. We cannot see a very real tree in front of us in the dark. Hence it is also reasonable -- and we argue, more reasonable -- to systematically vary the information contained in the light and let the ego - environment states corresponding to that information vary freely rather than the other way around as is now the practice. There would be no problem as to which to deliberately vary and which to let vary freely if simple or low-order optical and environmental structures were in one-to-one correspondence, but that they are not always so has been plaguing the study of perception since Euclid.

An example in which there is lack of correspondence between simple or low-order optical and environmental states is common in aviation: Two planes may be traveling at the same ground speed, but if one is flying very low, both the optical flow rate and the corresponding experience of egospeed will be fast, whereas if one is flying very high, both the optical flow rate and the corresponding experience of egospeed will be slow (Warren, Note 1). Hence, a study that systematically varied egospeed, but not optical flow rate, could miss the dependence of perceived egospeed on altitude. A study that included altitude as a second orthogonal factor might find a significant interaction between egospeed and altitude, but unless it went beyond the environmental factors to the relevant optical factor, it could not explain the interaction. There are two lessons to be learned from this example: One is concerned with the number of factors to include in an experiment and is discused in the next section. The other lesson is that the finding of a functional relationship between an environmental condition and perception is not enough, for we must also learn what the information "linking" the two is. Unfortunately, the optical conditions, especially the optical invariants, tend to be ignored.

Critique of Assumption 2

The second assumption of the standard approach may be characterized as implying that the factors chosen for an orthogonal design may be so chosen and so presented as to avoid the effects of any confounding factors either by elimination or deliberate control of all possible confounds. Our point here is that this situation, however desirable for elegance of design and ease of interpretaion, is in general inherently unattainable in experiments utilizing scenes of sufficient ecological validity to be of interest in aviation research. In general, there will exist at least one, and often many, identifiable factors, in addition to the specified set of orthogonal experimental factors, which will stand in a non-orthogonal relationship to them. In other words, there will always exist confounding factors whose effects cannot be controlled or eliminated by the experimenter, because the factors are inherently tied given the environmental constraints.

Where the inherent confounding exists, the very notion of confounding must be reinterpreted. We will attempt a reinterpretation and try to specify the conditions under which aviation research leads to non-standard analysis.

The reason for the inherent confounding of experimental factors is that each experimental factor (excluding non-visual factors such as replications and flying experience) corresponds to some characteristic or descriptive parameter of the visual scene, whereas the number of degrees of freedom available for distribution among the scene parameters is smaller than the number of scene parameters that must assume values. One consequence of the shortage of degrees of freedom is that an experimenter may manipulate or specify the values of only a small subset of scene parameters; the values of all the other unavoidably co-existing scene parameters are then forced or determined once the values of the initial subset are assigned. The experimenter's problems are further exacerbated since there is not complete latitude in choosing which combination of scene parameters may be assigned to the degree of freedom consuming subset. This may be best explained by identifying the scene parameters and their interrelationships:

Scene parameter degrees of freedom. A complete description of an egomotion scene includes a specification of the environment and the orientation of the "window" through which an observer views the world. In addition, the following must be specified:

1. The path slope. The specification of the path slope consumes one degree of freedom.

2. Speed of travel. Speed of travel may refer to the path speed or to its components, descent rate and forward velocity. But assignment of values to these three parameters is constrained since they are related by the Pythagorean theorem: Path speed is the square root of the sum of the squares of descent rate and forward velocity. Another constraint is that descent rate and forward velocity are functionally related by the prior selection of a path slope since path slope is equal to the ratio of descent rate to forward velocity. These two constraints mean that there is only one degree of freedom for selecting among the three parameters of path speed, forward velocity and descent rate.

3. Initial position. The initial position of an observer in an egomotion scene consumes one additional degree of freedom. Position is fixed once one of the three position paramters of path distance to the touchdown point, ground distance to the touchdown point, or initial altitude is assigned a value. This is because path distance, on a rectilinear path, is related to the ground distance and the altitude by the Pythagorean theorem: Path distance is the square root of the sum of the squares of the ground distance and the altitude. Another constraint comes from the prior selection of path slope since path slope, in rectilinear travel, is equal to the ratio of the altitude to the ground distance.

4. Initial acceleration. The acceleration aspect of travel also permits one degree of freedom for its determination in a manner entirely analogous to the cases of initial postion and initial speed. The three parameters of path acceleration, forward acceleration, and downward acceleration are determined once the value of one is chosen.

5. Ground texture size. Computer generated displays often use ground texture that is regular or stochasticaly so. The determination of the (average) texture unit size also consumes one degree of freedom.

Summary of degrees of freedom. The 11 scene parameters just descibed permit only five degrees of freedom for their selection.

Further restrictions. An experimenter is further constrained in that the five degrees of freedom may not be distributed freely. This is because certain combinations of variables are mathematically related and that relation cannot be broken. For example, since path slope is the ratio of descent rate to forward velocity, no experiment may orthogonally vary all three factors. This can be very frustrating to the researcher who wishes to determine the effects of these factors on flying performance. Another example is provided by the problem of determining the relative effects of the various variables that might affect the perception of change in altitude: No ecologically valid set of egomotion displays may simultaneously orthogonally combine the factors of descent rate scaled in meters, in eyeheights, in ground texture units, and the ratio of descent rate to forward velocity, since there are only three degrees of freedom available for these variables. But the expermenter's quandary is further deepened because the honorable techniques of setting one factor to a constant value or eliminating it are not applicable. All four factors must coexist, and due to their functional dependencies, one will always vary outside of the experimenter's control.

"Primary" and "secondary" independent variables. In an experiment, the factors that an experimenter chooses to manipulate are generally refered to as IVs and are here further specified as "primary" IVs. The factors that exist as a consequence of the mathematical relationships among the primary IVs are also true IVs in spite of the fact that they are not orthogonal to the primary IVs and that they assume their values as a function of their relationship with variables controlled directly by the experimenter. Thus, primary IVs correspond to the subset of scene parameters to which the experimenter has chosen to allocate the available degrees of freedom. The secondary IVs then correspond to the scene parameters not manipulated by the experimenter.

What is a primary IV in one experiment may become secondary in another experiment. For example, in one experiment, an experimenter may orthogonally cross descent rate and forward velocity as primary variables. Path slope is then determined by the ratio of descent rate to forward velocity and is a legitimate experimental factor although the experimenter did not assign its values directly. In another experiment, the experimenter might choose to orthogonally cross descent rate with path slope, letting forward velocity vary as neeeded. In this second experiment, path slope has become a primary IV and forward velocity a secondary IV. No member of a mathematically related set of factors is inherently primary or secondary despite the appearance of the equations specifying the relationship. Any equation may be rewritten so that any variable appears as a function of the others.

It is important to note that the choice of primary and secondary IVs refers only to activity by an experimenter and not to activity by a perceiver or perceptual system. The experimenter's activity is to affect the availabiliy of optical information by manipulating directly the levels and ranges of the primary scene parameters and indirectly the levels of the secondary scene parameters. The perceiver's or perceptual system's activity is to pick up and utilize information from the optic array. A perceiver also may act to bring an event and its information into being as in the case of making a landing approach. But, which optical invariants are functional optical invariants for a given perceptual system is determined, in part, by the information extraction (not merely transducing) characteristics of that system and not by what the experimenter does. The lesson here is that the information that a perceiver uses may not always be the information that an experimenter was primarily manipulating. Analysis of the performance data as a function of the secondary IVs may reveal the effectiveness of these sources in contrast to the possibly less useful (or unused) primary IVs. This possibility has implications for the tenability of the third assumption of the standard paradigm.

Critique of the Third Assumption

The ecological critique of the third assumption is simply that it is not sufficient to just report that an interaction exists between two or more variables. In a simple experiment in which all confounding effects are eliminated and especially when the the experimenter has no theoretical expectation of a mathematical relationship between two variables, it may be reasonable not to pursue an analysis beyond the determination of the regression equation for the variables and their interaction. This is because there is no reason to "create" a new variable to enter into the regression equation. But in the complex visual scenes of the type encountered in aviation research, there do exist secondary IVs as a consequence of the mathematical relationships among the primary or main IVs in a standard orthogonal design. The mathematical relationship often takes the form of a decidedly non-additive "interaction" of the primary factors such as their product or ratio. Thus, it might be possible to specify the exact form of how the factors interact. This is preferable to merely concluding that "some" interaction exists.

Toward a New Methodology

The traditional experimental method, with its insistence on controlling and excluding confounding factors, is too powerful a research tool to dismiss lightly. But the visual scenes used in aviation research do seem to preclude the total elimination of "confounding" factors, and we have seen that sometimes these so-called confounding factors are very much of interest. We would very much like to orthogonally cross certain sets of factors but unfortunately are logically prevented from doing so as in the case of descent rate, forward velocity, and path angle or in the case of the four scaling variations of descent rate, viz., descent rate scaled in meters, altitude, and ground units per second plus the descent rate as a fraction of forward velocity. Thus, experimental research in aviation psychology requires some modification of standard methodology. The following list is intended as a first attempt at grappling with the problems posed by aviation research.

Recommendation 1

Since the visual system extracts information from light, it is reasonable to include optical variables and not just environmental variables in the set of primary IVs. For example, global optical flow rate can be included in the primary set in lieu of or crossed with path speed.

Recommendation 2

Since there is good theoretical reason to expect much, if not all, optical information to take the form of optical invariants, especially invariant ratios, it is important to include several levels of the optical invariant in question and also to form each level of the invariant using different combinations of absolute environmental values. The inclusion of several levels of an optical invariant permits assessment of whether or not the optical invariant is a functional optical invariant. Three levels within a range optimized by preliminary experimentation will typically reveal the form of the functional relationship. For an optical invariant to be a functional optical invariant, performance must vary as the optical invariant is set to different values. For example, does ability to detect the point on the ground toward which one is flying vary as the the angular separation between the focus of expansion and the horizon, an optical invariant under rectilinear egomotion, is set to different values? The forming of each level of the invariant from several combinations of absolute environmental values is for the purpose of enabling the invariant to exist independent of particular absolute levels of the component variables. An invariant can exist over the change or transformation within an event and also between events whose absolute values differ. For example, Table 1 shows that a path slope of .10 is common to three different flight paths having, in arbitrary units, descent rate / forward velocity pairings of 1/10, 2/20, and 4/40 respectively. If only one combination of absolute values were used, it would not be possible to attribute the results to the ratio or to the absolute values.

Table 1

Path slope as a function of
descent rate and forward velocity.
(arbitrary velocity units)

		Descent Rate		
		1	2	4
Forward Velocity	10	.10	.20	.40
	20	.05	.10	.20
	40	.025	.05	.10

Recommendation 3

Make all known secondary IVs explicit. Generally, experimenters report only the primary IVs that they used in an experiment and these are generally environmental rather than optical variables. But the secondary IVs are nevertheless present. Sometimes it is possible from the experimental report to determine some of the secondary IVs, but this is not always possible and poses unnecessary problems for readers. More frustrating is the all too common problem that, whether or not the secondary IVs are reported, the results, such as means, for these variables are impossible to compute from results summarized over levels of a variable. (A table of means for each cell in the design would solve this problem.) Results for the secondary IVs might actually be more impressive than those for the primary IVs and thus should be reported.

Recommendation 4

Recommendation 4 follows immediately from Recommendation 3: the statistical analyses should be extended to include the secondary IVs. Since the secondary IVs are generally non-orthogonal to the primary set, this means that multiple regression and stepwise multiple regression would be appropriate. Since multiple regression can be cumbersome, it would be useful to have a simple way to evaluate the secondary IVs taken one at a time. The following techniques are presented only as working suggestions, and since the statistical procedures need further evaluation, the results obtained should also only be treated as suggestive.

One technique to simply assess a secondary IV is to ignore all other variables and perform a one-way analysis of variance on the data. The number of levels of the secondary IV will be determined by the number and spacing of levels of the primary IVs "interacting" to produce the secondary IV. The nature of the combinations is such that the data for each level of the secondary IV represent a pooling of the data from one or more of the "primary" data cells produced by the orthogonal crossing of the primary IVs. The number of primary data cells that are pooled into one level of the secondary IV will, in general, not be equal, and hence the number of data points per level of the secondary IV will also not be equal. For example, assume Table 1 represents the design of a simple experiment with descent rate and forward velocity as primary IVs. In addition to a standard analysis, the data may also be analyzed for the effects of path slope as a secondary IV. Notice that this particular spacing of the three levels each of the primary IVs yeilds five levels of path slope. In particular, a path slope of .10 is formed by three different crossings of the primary variables whereas a path slope of .40 results from only one crossing. Assuming equal numbers of data points per primary cell, then there are three times as many data points at the .10 level of path slope as there are at the .40 level since the data for the .10 level come from the pooling of three primary cells whereas the data for the .40 level come from only one primary cell.

There are two reasons for arguing that a one-way analysis of variance is appropriate for the assessment of a secondary IV. One reason is that one-way ANOVA is well suited for and unambiguous with respect to the unequal "$\underline{n}$" problem that arises from the pooling of different numbers of primary cells to yield the levels of the secondary IVs. The problem of unequal "$\underline{n}$" within the context of complex ANOVA is, of course, notorious. Another reason for suggesting the one-way ANOVA is that the ratio of the between-groups sum of squares to the total sum of squares is equal to the coefficient of curvilinear determination (η^2) and the coefficient of multiple determination ($\underline{R}^2$). This ratio indicates the proportion of variance accounted for by all the statistical information in the secondary IV and thus is an index of the total strength of the variable.

However, extreme caution must be used in interpreting the $\underline{R}^2$ produced by the above method. Its strength of using all the statistical information in the secondary IV is also its weakness. The $\underline{R}^2$ so obtained is that $\underline{R}^2$ obtained by a regression line fitted perfectly through the means of the secondary IV. That regression equation is a power polynomial of degree equal to the number of levels of the variable less one. The problem with such a regression line is that it accounts for too much: every kink and outlyer in the means is fitted. A "true" functional relationship, on the other hand, generally implies a smooth trend line through the means.

The determination of a smooth trend line depends on the particular data. Although mathematical curve fitting and trend analysis procedures are left to other sources, we emphasize that a visual inspection of the graph of means is the best first step, and that the trend equation need not be a power polynomial: Power or logarithmic functions are often more common and interpretable with respect to psychological theory. Whatever the trend equation, the proportion of variance accounted for by that equation is given by the ratio of the sum of squares due to trend to the total sum of squares (*SS*trend / *SS*total). This ratio, asssuming a judiciously chosen trend line, will give a more reasonable estimate of the proportion of variance accounted for by the true relationship between the performance measure and the secondary IV, all other factors excluded.

In particular, if the means appear to have a logarithmic trend (which implies that constant increments in performance correspond to constant proportionate increases in the secondary IV), then the "honest" proportion of variance accounted for by the secondary IV is directly provided by the coefficient of linear determination (r^2) between the performance measure and the logarithm of the secondary IV.

Recommendation 5

In a two-factor balanced equal-"*n*" orthogonal design, the between cell sum of squares (*SS*cell) is equal to the sum of the sums of squares for each factor and their interaction: *SS*cell = *SS*a + *SS*b + *SS*ab. In this sense, the *SS*cell exhausts all the statistical information available in the primary IVs. This statement is also true for designs involving more than two factors with appropriate inclusion of all relevant main effects and interactions. The *SS*cell may be formed for all the Primary IVs or for just a select subset. A subset of the primary IVs might be selected when, for example, an optical invariant can be formed using only some of the primary IVs in an experiment. The ratio *SS*cell/*SS*total is the total proportion of variance in the performance measure accounted for by all the statistical information in the primary IVs and their interactions. We suggest that this ratio can then serve as a reference or benchmark level against which the strength of any secondary IV may be compared.

An index of how well a particular secondary IV (SIV) accounts for the data as compared to the (relevant) primary IVs is given by:

$$(SS\text{siv}/SS\text{total})/(SS\text{cell}/SS\text{total}) = (SS\text{siv}/SS\text{cell})$$

But, as was just argued (in Recommendation 4 and letting the one-way ANOVA *SS*between there equal the *SS*siv here), *SS*siv is too strong a measure and can be artificially be made equal to *SS*cell by any artificial function that results in as many levels of the SIV as there are primary cells. A more "honest" procedure is to use the proportion of variance accounted for by a smooth regression line through the means of the SIV, viz., *SS*trend/*SS*total. An index of how well the smoothed SIV function compares to the primary variables is given by:

$$(SS\text{trend}/SS\text{total}) / (SS\text{cell}/SS\text{total}) = (SS\text{trend}/SS\text{cell})$$

As a special case, if a logarithmic trend is manifest, the r^2 for the log of the SIV may be used directly:

$$r^2/(SS\text{cell}/SS\text{total})$$

Notice that no SIV, however defined, can account for more variance than that accounted for by the primary IVs from which it is formed. But the SIV does represent a legitimate alternative interpretation of the data and may account for more variance than any single primary IV or interaction.

The above technique needs further study. For example, the proportion of variance accounted for by the SIV, either from the one-way ANOVA or the trend analysis, is obtained from a set of data with unequal "_n_"s for the SIV levels. Whether or not this instance of unequal "_n_" affects the analysis in any material way remains to be determined. Another area to be investigated is the use of _SS_cell for comparison purposes. In an unequal "_n_" design, it is not generally true that the _SS_cell equals the sum of the sums of squares of the main effects plus their interactions. What an experimenter should do in such a situation is not yet totally clear. Hence, the above procedures are offered as tentative suggestions, but nevertheless some method must be developed to enable assessment of the effects of the SIVs. The suggested procedures do show promise. They are easy to use and to interpret and there is reason to believe that if they are not precisely on target, they are not far off. At the very least, they serve a heuristic purpose in choosing primary IVs for subsequent experiments.

Raw data vs. means. So far the discussion has assumed that the entire data set was being analyzed. The variance not due to cells, (_SS_total - _SS_cell)/(_SS_total), includes the effects of "pure error", individual differences, practice, etc. It can be argued that it is unfair to expect a theory to account for such variance when evaluating a model (Cohen & Cohen, 1975, p. 249). A simple way to exclude practice and observer effects is to perform a regression analysis on only the means of the variables under study. For example, the r^2 between the means of a performance measure and the log of the SIV indicates how well a logarithmic function fits the means, all practice and observer effects excluded. Such an r^2 , by itself, can be comparable, if not identical, with the ratio r^2/(_SS_cells/_SS_total) defined earlier for the entire data set. The r^2 obtained using only the means will, of course, have many fewer degrees of freedom associated with it than the r^2 for the entire raw data set and this may affect the significance level.

Conclusions

The main point of this paper is that the visual displays encountered in aviation psychology research unavoidably make available optical information in addition to the information they are designed to present. Hence, experiments designed to assess the utilization of different sources of information in aviation are subject to alternate interpretaion. The experimenter is then faced with the problem of determining which of several (partially) redundant sources of information is actually responsible for a pilot's performance. These problems will become more evident and more formidable when the experimenter turns control of the environmental and optical variables over to the pilot in fully interactive flight situations, simulated or actual.

Although no solution yet exists, some statistical procedures are tentatively proposed to determine the relative strength of each factor. Whatever the fate of these particular proposals, some assessment procedure must be found that is applicable to experimental research in aviation. Paradoxically, the situation of the aviation experimenter is more akin to that of the non-experimental field researcher and hence, the multiple regression techniques developed for many-factor non-experimental data may prove useful.

Reference Note

Warren, R. Global optical bases for the perception of egospeed. Manuscript submitted for publication, 1981.

References

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N. J.: Erlbaum, 1975.

Gibson, E. J. Principles of perceptual learning and development. New York: Appelton-Century-Crofts, 1969.

Gibson, J. J. Motion picture testing and research. AAF Aviation Psychology Research Report No. 7. Washington: Government Printing Office, 1947.

Gibson, J. J. The ecological approach to visual perception. Boston: Houghton Mifflin, 1950.

Gibson, J. J., Olum, P., & Rosenblatt, F. Parralax and perspective during aircraft landings. American Journal of Psychology, 1955, 68, 372-385.

Owen, D. H., Warren, R., Jensen, R. S., Mangold, S. J., & Hettinger, L. J. Optical information for detecting loss in one's forward speed. Acta Psychologica, in press.

Shaw, R., & Bransford, J. Introduction: Psychological approaches to the problem of knowledge. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, N. J.: Erlbaum, 1977.

ACKNOWLEDGEMENTS

Research sponsored by the United States Air Force Systems Command, Office of Scientific Research under grants AFOSR-81-0078 and AFOSR-81-0108. We thank Rich Jagacinski for helpful discussions on curve fitting. Request reprints from Rik Warren, Psychology - O.S.U., 404C W. 17th Ave., Columbus, OH 43210.

FRACTIONAL RATES OF CHANGE
AS FUNCTIONAL OPTICAL INVARIANTS

Susan J. Mangold, Dean H. Owen, and Rik Warren

The Ohio State University

ABSTRACT

Two experiments were conducted to determine whether higher-order optical information is used to detect loss in forward velocity and loss in altitude. For the deceleration experiment, the lower-order environmental variables of deceleration rate, altitude, and initial forward velocity were combined to produce variables and invariants for initial optical flow rate, optical flow deceleration, and initial optical flow damping. For the descent experiment, descent rate, initial altitude, and forward velocity were varied to generate levels of optical flow rate, path angle, and relative rate of change in optical flow, density, and splay angle. Speed and accuracy of responding were constant when invariant values of fractional loss in speed or altitude were compared, in spite of the varying values of the lower-order variables contributing to the ratios. These results suggest that future research should investigate the psychological effectiveness of higher-order optical information for control of self motion, in contrast to the past emphasis on lower-order environmental variables.

INTRODUCTION

The use of visual information available outside the cockpit is critical to a variety of aircraft maneuvers, one example being the landing phase of the flight. With the trend toward increased use of flight simulators for both training and experimental purposes arises the need for determining specifically the nature of this information in order to ensure that it is accurately represented by the visual display simulating the external scene. Failure to supply this information may cause difficulties in interpreting experiments which use flight simulators.

Although some researchers have assessed pilots' sensitivity to visual information, the parameters manipulated have tended to be environmental variables. A common approach to aviation research involves the manipulation of lower-order variables, such as forward and vertical velocities, in order to assess resulting variations in performance. For example, Palmer and Petitt (1977) used computer-graphic night visual simulation in an attempt to determine why control of altitude during flare and touchdown is much more difficult in a simulator than in an airplane. Sensitivity to variations in descent rate prior to touchdown was measured by asking airline pilots to judge which of two scenes presented a faster sink rate. Height above the runway was also varied to assess its influence on detection of differences in sink rate. Implicit in this approach is the assumption that the visual system can directly make use of this absolute information (altitude and descent rate) scaled in meters or meters per second.

An alternative conception of visual information proposed by Gibson (e.g., 1966, 1979) maintains that the visual system is sensitive to higher-order information which specifies both the environment and the individual's movements through that environment. This information is body-scaled, that is, it is in a form which the perceiver can directly use to control movement without subjecting the information to processing. With respect to the manipulated variables of Palmer and Petitt, it can be argued from the ecological approach to perception that it is not descent rate in meters per second which is of consequence to the visual system but rather the magnitude of the change relative to the initial altitude at which the event began. The resulting ratio, descent rate as a percent (or proportion) of initial altitude, scales descent rate in terms of the observer's eyeheight and thus provides a higher-order optical variable (Warren, Note 2). If this is, in fact, a source of information to which the visual system is sensitive, performance should not vary with variations in the values of altitude and descent rate so long as the ratio itself does not change. If, on the other hand, it is the lower-order variables that are actually used, performance will not vary with variations in the higher-order ratio when the lower-order variable remains constant.

Because there are these two alternative conceptions of visual information, the experimental paradigm to be used in future investigations of visual information specifying self motion must be carefully selected in order to allow measurement of performance as a function of both of these descriptions of information. Failure to select such a paradigm can have the consequence of limiting experimental conclusions to those founded upon an inappropriate description. The possibility then arises that the visual system may not actually be sensitive to the lower-order variables the experimenter assumes are being manipulated. What is required is an approach which allows assessment of performance as a function of both the lower-order and higher-order variables.

The paradigm used in the two experiments presented here involves selecting values of lower-order variables so as to produce a matrix containing both higher-order invariants for which the levels of the lower-order variables differ but the ratios are equal, as well as higher-order variables for which the ratios are not equal. Three levels of each of the two lower-order variables produce a nine-celled matrix in which the higher-order variables are found on the negative diagonals and the five variable levels appear along the principle positive diagonal (see Table 1). The advantage of this paradigm is that performance can be assessed as a function of both the lower-order and higher-order variables. Consequently, it is possible to determine which type of information most adequately reflects performance. This paradigm was used in an experiment to identify sensitivity to information which specifies change in forward velocity (Owen, Warren, Jensen, Mangold, & Hettinger, in press). Information for change in altitude was investigated by Owen, Warren, and Mangold (Note 1).

Experiment 1: Information for Change in Forward Velocity

In the Owen et al. (in press) experiment, deceleration rate, initial forward velocity, and altitude were varied in order to produce variables and invariants of initial optical flow rate, optical flow deceleration, and initial optical flow damping, as shown in Table 1. Optical flow deceleration, the ratio of deceleration to altitude, is a scaling of deceleration in eyeheights. Eyeheight is defined as the observer's altitude above the surface of the ground. For example, a pilot flying level at 80 meters has an eyeheight of 80 meters. As shown in Table 1A, forward velocity can then be scaled in number of eyeheights per second, which provides a measure of global optical flow rate (see Warren, Note 2). The pilot at 80 meters who has a forward velocity of 80 meters per second is traveling at one eyeheight per second. Similarly, global optical flow deceleration can be specified as deceleration scaled in eyeheights, as shown in Table 1B. Unlike deceleration scaled in an arbitrary metric such as meters, global flow deceleration varies with altitude.

Table 1

Combinations of Environmental Variables to Produce Optical Variables and Invariants for the Deceleration Experiment

A. Initial Global Optical Flow Rate (eyeheights/sec)

	Initial Forward Velocity (m/sec)		
Altitude (m)	18	31.5	55.125
20	.900	1.575	2.756
35	.514	.900	1.575
61.25	.294	.514	.900

B. Global Optical Flow Deceleration (eyeheights/sec/sec)

	Deceleration Rate (m/sec/sec)		
Altitude (m)	2	3.5	6.125
20	.100	.175	.306
35	.057	.100	.175
61.25	.033	.057	.100

C. Global Optical Flow Damping (% eyeheight/sec)

Initial Forward	Deceleration Rate (m/sec/sec)		
Velocity (m/sec)	2	3.5	6.125
18	11.1	19.4	34.0
31.5	06.3	11.1	19.4
55.125	03.6	06.3	11.1

A more likely candidate for specifying loss in forward velocity is initial optical flow damping, obtained from the ratio of deceleration to initial forward velocity. Optical flow damping scales deceleration as a proportion of the initial forward velocity. It was hypothesized that performance should improve as optical flow damping increases due to the proportionately larger ratio of deceleration to forward velocity. Performance should not vary with variations in the lower-order values so long as the initial damping ratio itself does not change.

Method. The ground surface simulated consisted of random rectilinear blocks 10 m wide and from 5 m to 20 m in length. Four ground texture colors (light green, dark green, tan, and brown) were randomly assigned to the blocks so as to form blocks of various sizes and shapes. A special purpose visual scene generator produced 10-sec sequences representing self motion over the surface. An analog computer was programmed to control the scene generator in order to vary altitude, initial forward velocity, and deceleration. The scenes were videotaped and played back on a Sony video projector screen 150 cm wide and 112.5 cm high, providing a field of view of 34 deg by 26 deg when viewed from a distance of 2.43 m.

Three levels each of deceleration rate (2, 3.5, 6.125 m/sec/sec), altitude (20, 35, 61.25 m), and initial forward velocity (18, 31.5, 55.125 m/sec) were fully crossed to produce 27 deceleration scenes. In addition, 27 constant-velocity scenes were obtained by crossing the three levels of altitude with the three levels of forward velocity. The nine combinations were each presented three times. Each observer received three blocks of the resulting 54 deceleration and constant-velocity trials. Each block consisted of a different random order of the scenes, and all six possible block sequences were used an equal number of times.

The observer sat in a Link General Aviation Trainer-1 flight simulator positioned 1.67 m above the floor, a height equal to the horizon displayed on the screen. The observer viewed the visual scene and judged whether the scene represented constant velocity or deceleration. The response involved pressing one of two buttons and was followed by a confidence judgment on a three-point scale. Time from event onset to response was recorded from a millisecond timer without the observer's knowledge. Consequently, no stress was placed on speed of responding although observers were encouraged to respond during the 10-sec scene duration. No feedback concerning performance was provided during the experiment.

Forty-two male undergraduates served as observers in order to fulfill a course requirement. None of the observers claimed to have had prior experience in visual flight simulators.

Results. Analyses of variance were performed on all of the data to be discussed. With the exception of a non-significant effect of blocks (i.e., practice), in every case significance was obtained at the .0001 probability level.

Performance on deceleration trials was first evaluated as a function of the lower-order variables. Proportion errors increased as altitude increased, but this effect accounted for only .3% of the variance, using the η^2 (eta square) measure. Proportion errors decreased as deceleration increased, accounting for 12% of the variance. Faster forward velocities resulted in an increase in proportion errors (η^2 = 7.3%). Mean reaction time increased as altitude increased, but the effect was again small, accounting for only .9% of the variance. Faster forward velocities resulted in slower mean reaction times (η^2 = 8.6%), while faster decelerations were accompanied by faster mean reaction times (η^2 = 17.6%).

Optical variables were similarly analyzed with the η^2 measure. Optical flow rate failed to account for much of the variance, having an η^2 = 2.6% for proportion errors and 2.2% for mean reaction times. A stronger effect was found with optical flow deceleration, with 8% of the proportion-error and 13.2% of the mean reaction-time variance attributed to this variable. By far, the greatest variance can be credited to optical flow damping. For proportion errors, η^2 = 23.2%, while mean reaction time produced an η^2 of 26.4%.

Based upon the η^2 measure, there is little question that sensitivity to loss in forward speed is most closely related to optical flow damping. Figures 1A and 1B show proportion errors and mean reaction time as a function of forward velocity and deceleration rate. The solid horizontal lines, which connect the invariant damping ratios, reveal that almost constant levels of performance resulted in spite of variations in the lower-order variables contributing to the ratio. In addition, performance varied with variations in the values of optical damping. Together, these findings suggest that optical flow damping is an effective source of information used in detecting loss in forward velocity. Such a higher-order invariant is thus a functional invariant.

Experiment 2: Information for Loss in Altitude

The same paradigm was used by Owen et al. (Note 1) to investigate information which specifies change in altitude. Again, lower-order variables were selected in order to produce matrices of higher-order variables.

Method. Three levels each of initial altitude (20, 40, 80 m), forward velocity (18, 36, 72 m/sec), and descent rate (1.25, 2.50, 5.00 m/sec) were selected in order to obtain 27 descent trials. Each of the nine unique constant-altitude trials were presented three times. Table 2 displays the higher-order optical variables which result from crossing the lower-order environmental variables. Observers received only one block of trials and a separate randomization was developed for each observer with the constraint that no more than four trials of the same type, level or descent, could occur in a row. The task involved judging whether the scene represented level flight or descent, and the observer specified his decision by pressing one of two buttons and then made his confidence judgment. Twenty male undergraduates participated in the experiment for course credit.

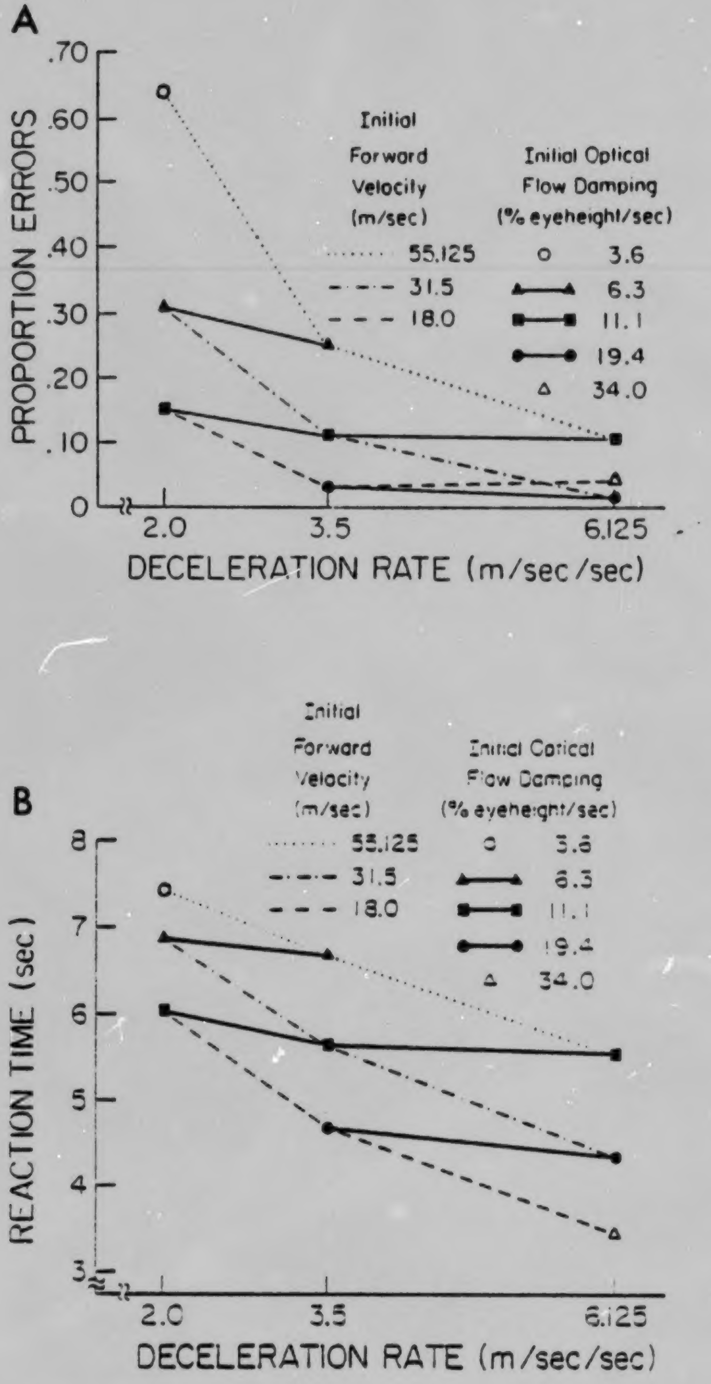


Figure 1. Mean proportion error (A) and mean reaction time (B) as a function of deceleration rate, initial forward velocity, and their ratio (initial global optical flow damping). Each point represents 378 observations (42 observers x 3 altitudes x 3 replications). Solid lines show error rates for invariant values of initial optical damping; broken lines are for different initial altitudes.

Table 2

Combinations of Environmental Variables to Produce Optical Variables and Invariants for the Descent Experiment

A. Initial Global Optical Flow Rate (eyeheights/sec)

Initial Altitude (m)	Forward Velocity (m/sec) 18	36	72
20	.900	1.800	3.600
40	.450	.900	1.800
80	.225	.450	.900

B. Path Angle (%)[1]

Descent Rate (m/sec)	Forward Velocity (m/sec) 18	36	72
1.25	6.94	3.47	1.74
2.50	13.89	6.94	3.47
5.00	27.78	13.89	6.94

C. Relative Rate of Change in Optical Flow, Density, and Splay Angle (% eyeheight/sec)

Initial Altitude (m)	Descent Rate (m/sec) 1.25	2.50	5.00
20	6.25	12.50	25.00
40	3.13	6.25	12.50
80	1.56	3.13	6.25

[1] Path angle is an optically available angle since it is represented in the array by the angular separation between the horizon and the focus of expansion.

Results. Analyses of variance showed significant main effects of altitude for both proportion errors (η^2 = 2.2%) and mean reaction time (η^2 = 13.1%). Descent rate also was significant for both proportion errors and mean reaction time (η^2 of 3.9% and 17.3% respectively). The third lower-order variable, forward velocity, failed to reach significance for either measure.

Initial optical flow rate did not account for much of the variance, having an η^2 of 4.8% for mean reaction time and .9% for proportion errors. A somewhat stronger effect was obtained with the higher-order invariant of path slope, the ratio of descent rate to forward velocity. Mean reaction time for this variable accounted for 11.3% of the variance and 3.8% of the proportion-error variance. The greatest variance is attributed to descent rate as a percent of initial altitude. The mean reaction-time η^2 was 30.5% and the proportion-error η^2 was 7.4%. In Figure 2, the solid horizontal lines demonstrate that performance as a function of initial fractional loss in altitude is fairly constant when error rates were low.

Under the conditions of the descent experiment three different but redundant sources of optical information which specify descent rate were available to the observer for detection of loss in altitude.

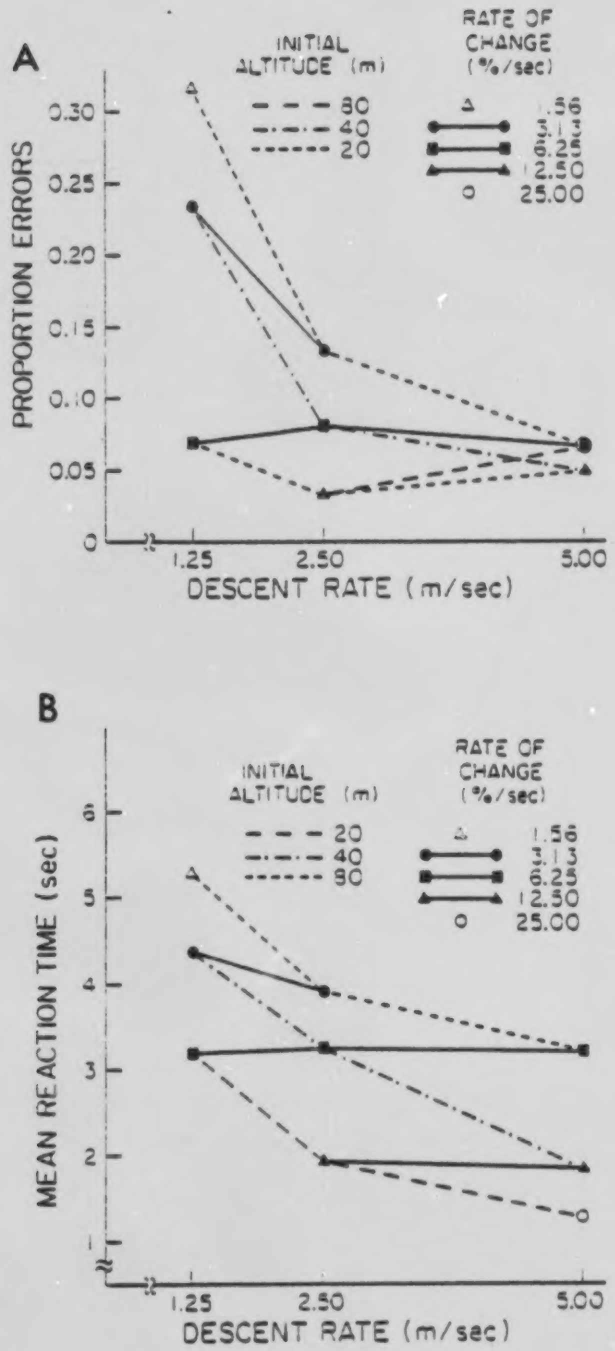


Figure 2. Mean proportion error (A) and mean reaction time (B) for the nine combinations of descent rate and initial altitude. Each point represents 60 observations (20 observers x 3 forward velocities). Solid lines show error rates for invariant values of initial relative rate of optical change; broken lines are for different initial altitudes.

Relative rates of change in optical flow, optical density, and optical splay (Warren, Note 2) were identical. On descent trials, the optical flow rate is constantly increasing due to the decreasing eyeheight, producing an accelerating function which describes optical flow acceleration. Optical density, scaled in ground units per eyeheight, is defined as the number of ground texture elements spanned by a single eyeheight. Decrease in altitude thus is accompanied by decreased optical density. Optical splay rate is the angular increase in parallel surface lines projecting to the vanishing point as altitude decreases. Experiments are currently in progress which will isolate the three types of information.

Discussion

The Owen et al. (in press, Note 1) experiments suggest that relative or fractional rates of optical change are useful information for detecting change in self motion. Optical flow damping, or fractional loss in forward velocity, was shown to be a source of psychologically effective information in detecting loss in forward velocity. Sensitivity to change in altitude, on the other hand, is a function of fractional loss in altitude. In both cases, accuracy and speed of responding varied with the magnitude of the change, either change in forward velocity or altitude, relative to the initial conditions. Both types of higher-order information were demonstrated to be functional invariants in that they each met the criteria that performance must remain fairly constant so long as the higher-order ratio is invariant and vary in an orderly fashion when the ratio changes.

These results complement the experimental literature on looming and time-to-collision. For example, Lee (1976) has proposed that both the time at which braking is begun and the magnitude of the braking action in automobile driving are dependent upon rate of magnification in optical size of the vehicle directly ahead. The inverse of the relative rate of optical magnification is time-to-collision and thus provides a potential source of information about the continuation of an event even if the driver has no access to information concerning velocities of the two vehicles and the distance separating them. Lee's theory is founded upon Schiff's (1965) demonstrations that optical magnification is a higher-order source of information for imminent collision of an object with one's self, and Schiff's results support the contention that it is percent per second change which is directly used by the visual system.

Overall, these results demonstrate the need for testing of alternative candidates for information which may be more appropriate for the visual system. Use of this paradigm is not limited to comparisons of lower-order and optical information, however. The paradigm has the potential of permitting evaluation of observer sensitivity to different types of optical information which ordinarily are available under normal conditions. Such an approach can be used to identify which of the three types of information is most useful in detecting change in altitude.

Experiments such as the two described here provide a preliminary approach to the issue of determining the visual information that is useful in guiding self motion. Information available to the observer is manipulated by the experimenter in order to assess the resulting variations in performance. A next step in the investigation involves giving the observer control over this information in interactive experiments. The observer assumes control of the simulator in order to perform specific maneuvers. Modification of the simulated flight path automatically brings about changes in the values of both the environmental states and optical information. In this case, the information produced by the observer becomes a dependent variable.

Together, the two types of experiments converge on the identification of visual information used in guiding self motion. The consequence is that simulator designers will know which types of information must be represented in the visual display in order to maximize transfer of training and transfer of experimental conclusions to the operational setting. Sensitivity to relevant optical variables and invariants is the criterion for comparing both hardware configurations and training conditions.

Reference Notes

1. Owen, D., Warren, R., & Mangold, S. J. Optical information for detecting loss in one's own altitude. Manuscript submitted for publication, 1981.

2. Warren, R. Global optical bases for the perception of change in altitude during egomotion. In preparation.

References

Gibson, J. J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.

Gibson, J. J. *The ecological approach to visual perception*. Boston: Houghton Mifflin, 1979.

Lee, D. N. A theory of visual control of braking based on information about time-to-collision. *Perception*, 1976, *5*, 437-459.

Owen, D. H., Warren, R., Jensen, R. S., Mangold, S. J., & Hettinger, L. L. Optical information for detecting loss in one's own forward speed. *Acta Psychologica*, in press.

Palmer, E. A., & Petitt, J. Difference thresholds for judgments of sink rate during the flare. *Journal of Aircraft*, 1977, *14*, 540-543.

Schiff, W. Perception of impending collision: A study of visually directed avoidant behavior. *Psychological Monographs*, 1965, *79* (Whole No. 604).

ACKNOWLEDGEMENT

This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Contract Number F49620-79-C-0070 and Grant Number AFOSR-81-0078.

VISUAL SCANNING BEHAVIOR AND PILOT WORKLOAD

R. L. HARRIS, SR.#; J. R. TOLE*; A. T. STEPHENS*; A. R. EPHRATH**

# Automation Research Branch, NASA Langley Research Center

* Biomedical Engineering Center for Clinical Instrumentation
Massachusetts Institute of Technology

** Dept. of Aero. & Astro., M.I.T., and Bell Laboratories

ABSTRACT

This paper describes an experimental paradigm and a set of results which demonstrate a relationship between the level of performance on a skilled man-machine control task, the skill of the operator, the level of mental difficulty induced by an additional task imposed on the basic control task, and visual scanning performance. During a constant, simulated piloting task, visual scanning of instruments was found to vary as a function of the level of difficulty of a verbal mental loading task. The average dwell time of each fixation on the pilot's primary instrument increased as a function of the estimated skill level of the pilots, with novices being affected by the loading task much more than the experts. The results suggest that visual scanning of instruments in a controlled task may be an indicator of both workload and skill.

INTRODUCTION

A persistent difficulty in the design and evaluation of man-machine systems is the inability to accurately measure operator workload under various conditions. Some of this difficulty is due to an inability to define clearly what is meant by workload. So far, a single definition of workload has not been developed. This is not surprising since the topic is heavily related to the study of human behavior which itself is not an exact science. The term workload has been constructed to be descriptive of the difficulty of performing a task. One must usually rely on various descriptions of workload including verbal and graphical analogies. There are at least two components to workload: mental and physical. A list of some of the aspects of these two types is illustrated in figure 1.

The desire to measure workload is usually motivated by the need to predict situations in which operator performance will decline. The reasons for this need are evident: if the operator has too many tasks to accomplish in too short a time, the performance on all or some of the tasks may be diminished. The same may be true if the operator has too few tasks to perform and allows his attention to wane. For example, recent experiments (1) suggest that a general aviation pilot flying a simulator equipped with an autopilot has a decreased ability to detect his own blunders as the sophistication of the autopilot increases. It should also be noted that another important potential cause of performance decrement is the occurrence of an extremely rare or novel event or series of events. Pilot training methods attempt to take some of the more common of these rare events into account by having the pilot practice procedures for dealing with problems such as engine out, stall, loss of one or more instruments, etc. It may be the unusual failures which have never been seen before which represent the greatest difficulty since they may cause the pilot to focus his attention too narrowly, perhaps forgetting about his primary piloting task at a critical moment.

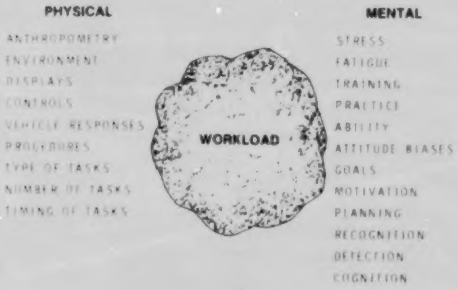


Figure 1.- Aspects of workload

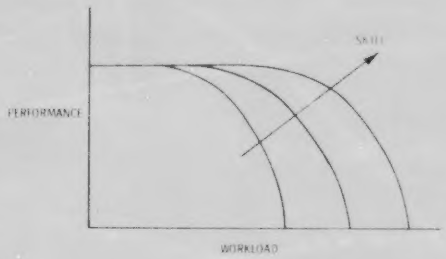


Figure 2.- Theoretical relationship of performance, workload, and skill

One of the important human factors questions in the cockpit is how to specify the procedures, display, tasks, etc. in such a way that severe over or underloading of the pilot will not occur in any of the anticipated circumstances. Thus, if one has a choice of 5 different procedures for accomplishing some task, it would be quite useful to compare the relative difficulty of the procedures and the effects of various perturbations or external disturbances on each procedure. Were a quantitative comparison possible, the selection of the "best" candidate procedure (or display format, etc.) might be greatly facilitated. Such measures could also be used to compare what is currently in use with an alternate approach.

Since the goal of workload measurement is the prediction of performance, it is often suggested that performance is the parameter which should be measured as the workload conditions are varied. Certain performance criteria may be set and when the pilot cannot meet them the level of loading may be judged to be too high. Such a technique assumes that performance varies in a consistent fashion with workload and skill. That is, for this approach to be generally useful, all pilots should experience about the same performance decrement for the same increase in workload. Experience suggests that this is not the case however. The point is that in situations such as piloting, where performing manual dexterity and verbal or mental activities simultaneously are especially important, preformance of a skilled operator may not show a great decrement until the workload is severe, and then a precipitous decline in performance may occur.

Figure 2 is a graphical statement of our hypothesis for the relationships between workload, performance, and skill. This hypothesis is specifically directed to the high workload situation. Performance may also decrease at low workloads. This figure does show that performance remains constant over a range of workloads regardless of skill level. However, the more skilled operator can maintain that performance level at higher workloads than the less skilled. In an attempt to confirm this hypothesis, we have been exploring these relationships by examining the behavior of aircraft pilots under varying task difficulty. In the work described here, we are concerned with the variation of a dependent variable, the visual scanning of instruments, as a function of skill level, inherent task difficulty, and the difficulty of an additional verbal mental loading task. We will explore how the timing of fixations on various instruments varies as a function of verbal task difficulty and the skill level of the pilot. We will discuss the implications in the results for the assessment of skill level in a task which requires skilled visual performance under varying task difficulty and in the evaluation of learning behavior in this type of task.

DESCRIPTION OF EXPERIMENTAL PROCEDURES

These experiments are concerned with relationships between "steady-state" levels of the various independent parameters: piloting performance, skill, and workload. The approach was to attempt to demonstrate whether consistent steady state effects of a constant mental loading condition could be observed. Thus, the piloting task and verbal mental loading task were held constant for a period sufficiently long enough to collect the data to evaluate the average effects of these conditions. A run length of 10 minutes was chosen as an estimate of the minimum amount of time required to provide a sufficient number of fixations to satisfy the assumption of steady state conditions. The piloting task chosen was a precision straight and level path with zero degree glide slope and live localizer with constant sensitivities on the needle movements while maintaining a constant heading and airspeed. In order to force some pilot vigilance on this task, a low level of turbulence was also introduced for each run.

A desktop general aviation instrument flight simulator was used to simulate these flight maneuvers. This simulator is shown in figure 3. Pilot lookpoint on seven instruments (Attitude Indicator, Directional Gyro, Altimeter, Vertical Speed Indicator, Airspeed, Turn and Bank, and Glide Slope/Localizer) was measured and recorded by an oculometer described below.

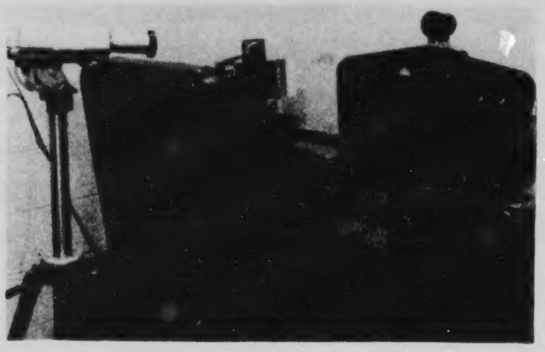


Figure 3.- Laboratory equipment

The mental loading task was chosen so as not to directly interfere with the visual scanning of the pilot (i.e. the task would not require the pilot to look away from the instruments in order to accomplish the task) while providing constant mental loading during the maneuver. This was accomplished by having the pilots respond verbally to a series of evenly spaced three-number sequences (4) presented to them by a tape recorder. The pilot was told that he must respond to each three-number sequence as either "plus" or "minus" (up or down respectively on a rocker switch) according to the algorithm: first number largest, second number smallest = "plus" (e.g. 5-2-4), first number smallest, third number largest = "plus" (e.g. 1-5-7), all other sequence combinations are "minus" (e.g. 9-5-1). The numbers were prerecorded at 4 second and 2 second intervals between sequences. These spacing intervals were determined empirically to create heavy and severe additional mental loading respectively. The pilot was instructed to give the number task a priority equal to that of the piloting task (as if the verbal task represented a constant rate of radio communication).

Pilot lookpoint was measured using a Honeywell oculometer system which has been substantially modified by NASA Langley Research Center (3). This device is non-invasive and allows the experimenter to determine the time course of eye fixations on instruments employed by the pilot and the dwell time of each fixation to the nearest 1/30 second. Starting with this information, dwell time histograms for each instrument for each loading condition could be plotted as discussed below. These histograms are a plot of the number of fixation counts which fall into bins of specified time duration during a run. In these figures, the bin size is 0.066 seconds and the range of the time axis is 0 to 5 seconds. The instrument name for each histogram is located on the left side of the figure.

Six subjects, varying in skill level from non-pilot to a highly experienced NASA test pilot participated in these experiments. The subject numbers and approximate skill level are listed below.

Subject Number	Skill Level
4	Highly skilled NASA test pilot
11	Highly skilled general aviation instructor pilot, NASA employee experienced in simulation experiments
9	General aviation pilot, current only in simulators
5	General aviation pilot, current in airplanes
10	Student pilot
7	Non-pilot

Subjects 5, 7, 9, and 10 are referred to collectively as "novice" pilots in this paper. Subjects 9 and 11 were the only pilots with any previous experience on this particular simulator. The subjects were allowed to practice the flying maneuver and verbal mental task until they felt comfortable with the situation.

RESULTS

Dwell Time Histograms

Perhaps the most striking effect observed in these experiments is the effect of the verbal loading task on the dwell time histograms of individual instruments for a given maneuver. In the four novice subjects, the dwell time on the primary instrument (the instrument which registered the highest total number of dwell counts) became progressively weighted toward extremely long dwells as the verbal task difficulty increased. Figure 4 shows the dwell time histograms for pilot 5 on the attitude indicator, directional gyro, glide slope/localizer, and vertical speed indicator. Note that for the no mental loading case, the dwell histogram on the attitude indicator has a fairly standard shape (2). When numbers are added to the piloting task, the dwell becomes longer and the peak of the histogram at 1/2 second begins to disappear. The effect is even more dramatic for the 2-second interval case; the entire distribution is skewed toward extremely long dwells on attitude as the pilot apparently begins to "stare" more and more at this instrument. Similar effects are seen for the other novice pilots.

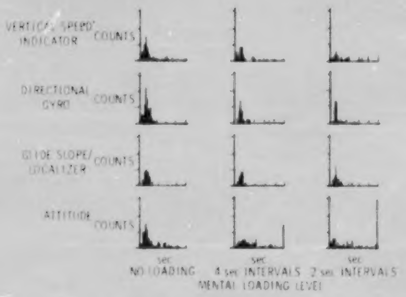


Figure 4.- Dwell histograms for pilot 5

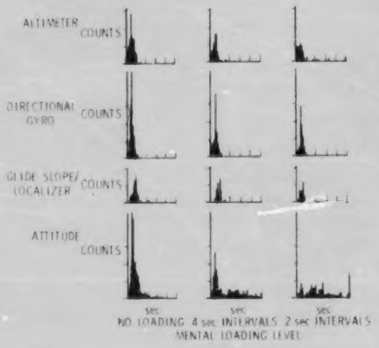


Figure 5.- Dwell histograms for pilot 11

However, an interesting difference occurs for subject 7, the non-pilot. This subject had no previous piloting experience and was only given enough practice to allow him to stay nominally on course during the precision straight and level maneuver. This subject adopted the glide slope/localizer as the primary instrument apparently in an effort to accomplish the precision task by keeping the needles centered. Even though the subject adopts the inappropriate instrument to accomplish the piloting task, the dwells on this instrument are affected in a manner similar to those on the attitude instrument for the more experienced novice subjects.

The visual scanning behavior of the two subjects with higher levels of skill was also affected by the verbal loading although to a much lesser degree than for the novice pilots. Figure 5 shows the dwell time histogram for subject 11, who had the next to highest skill level, and was somewhat more affected than the test pilot, especially at the highest loading level. Subject 11 uses a large number of short dwells on the attitude indicator under the no loading case. When the mental loading task is introduced at 4-second intervals, his distribution is shifted to somewhat longer dwells. However, there is still a very significant peak at around 1/2 second. The actual shift in dwell times is not as large as that seen in the novice pilot's histograms, even though there appears to be a large change due to the reduction in magnitude of the histogram peak. Even at the highest mental loading level, the shift to longer dwells is not as severe as it is in the less skilled pilots.

PILOT	MENTAL LOADING LEVEL[a]		
	NO LOADING	4 sec INTERVALS	2 sec INTERVALS
4	0	0.6	3.7
11	0	2.0	7.3
9	0	6.8	8.5
5	0	8.6	20.1
10	0	19.8	23.4
7[b]	0	6.9	13.2

a 0 MENTAL LOADING LEVEL SUBTRACTED OUT

b NONPILOT

Table 2. Percent of primary instrument dwells greater than 5 seconds

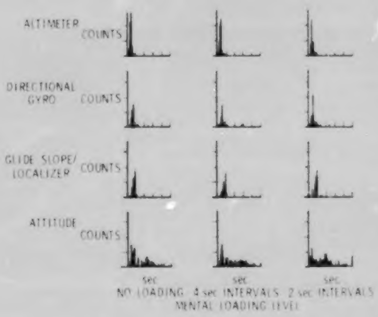


Figure 6. Dwell histograms for pilot 4

The shift to longer dwells may also be demonstrated by looking at the percentage change from the no loading case in the number of dwells on the primary instrument that are 5 seconds or longer in duration as the mental workload is changed. The raw counts of such dwell are shown as the last element in the histograms. Table II shows the percentage change from the no loading case for each pilot. The percentage of dwells is seen to increase with decreasing skill level. This holds for all subjects except subject 7, the non-pilot. It should be pointed out, however, that subject 7 used a different primary instrument from the rest of the pilots and, therefore, had a completely different basic scan pattern from the other pilots. This fact may not allow direct comparison of the results from subject 7 with the other subjects. This is not a cause for concern since the results from all of the pilot subjects seem to be consistent and, therefore, any conclusions drawn from their results should be applicable to other pilots.

The dwell time characteristics on secondary instruments were affected most in the novice subjects, as may be seen in figure 4. This is typical of all novice subjects. The secondary instrument dwells are seen to change in a different manner than the primary instrument dwells. As opposed to the shift to longer dwells, as in the case for the primary instruments, the effect of mental loading is to decrease the number of looks at secondary instruments, one example of a phenomenon know as load shedding. The shape of some of the histograms changes under varying loading conditions. Pilot 4 was the only subject whose dwell time histograms on secondary instruments were not affected by mental loading (figure 6). Subject 11 (figure 5) appears to exhibit some load shedding, primarily on the altimeter and vertical speed indicator.

Fixation Sequences

We are also interested in examining whether pilots develop a scan pattern or patterns during the constant flying maneuver in our experimental paradigm. Assuming that such patterns might exist, it appeared of interest to determine whether they might be altered by the addition of mental loading. The results from one method of studying this question is presented below.

If the dwell times on individual instruments are ignored, an ordered list of instrument fixations may be developed for each pilot for the various mental loading conditions. These lists may be broken up into smaller segments (or sequences) of various lengths for easier analysis. Each different sequence may be considered as a component of the overall scan pattern. One may hypothesize that the sequences which occur most frequently during the maneuver are those of most importance to the pilot and ones which might indicate an ordered scan pattern.

Examination of the results indicated that sequences of four instrument fixations were the longest for which there was a significant amount of repetition during a run, hence sequences of length four were chosen for analysis. The number of times each four instrument sequence occurred during a 10 minute run was obtained as was the total number of sequences of length four in the run. From this data, the percentage of occurrence was calculated for each observed sequence. For example there might be 800 sequences of length four in 10 minutes. If the sequence, ATT-DG-ALT-DG, occurs 40 times during the run, its percentage of occurrence would be 40/800 X 100 percent = 5 percent. In this fashion, the percentage of occurrence of all length four sequences in the no mental loading case was determined for each pilot. The 10 sequences which occurred most frequently were arbitrarily chosen as indicators of the scan patterns normally used by various pilots. The manner in which the percentage occurrences for these 10 sequences change as a function of mental loading is shown for two subjects in figures 7 and 8. Figure 9 plots the sum of these percentages across mental loading conditions for all subjects. It is important to note that the sequences used as the basis for calculation for all conditions are the ten most frequent for the no mental loading case. Each line beginning at the no mental loading case and ending at the 2-second interval case represents the same sequence.

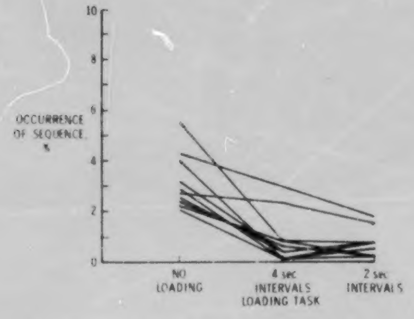


Figure 7.- Percent occurrence of sequences for pilot 5

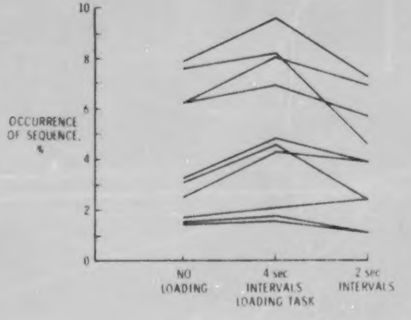


Figure 8.- Percent occurrence of sequences for pilot 11

Several interesting observations may be made by comparing the plots of the skilled pilots with those of the novice subjects. A difference may be seen between the two groups in the percentage of occurrence of the most often used sequences. The first ten sequences used by the skilled pilots comprise over 50 percent of their scan pattern (see figure 9). The usage of these ten sequences is relatively constant with changes in mental loading suggesting that the patterns are not disturbed by the mental loading task. This finding is certainly in keeping with the intuitive development presented in the introduction which suggested that it should be difficult to interfere with a skilled subject performing a task.

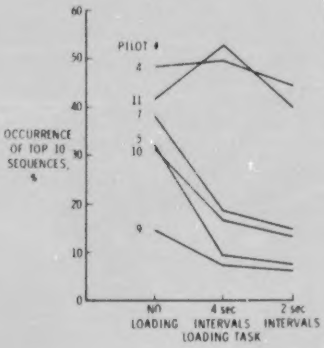


Figure 9.- Percent occurences of ten most frequent sequences for all pilots

The novice pilots' results differ in several respects from those of the skilled subjects, however. The ten most frequently used sequences in the no loading case occupy much smaller percentages of the total scan than do those of the skilled pilots. This suggests the novices' sums are more random than those of the skilled subjects, even without the imposition of an additional task.

The novice subjects also show a consistent decrease in the percentage occurrence of the ten sequences as the mental workload is increased. This decrease may be the result of either the equalization of the number of occurrences of each sequence in the run (i.e. a trend to randomization) or a change to a different set of sequences from those used in the no loading case (i.e. a change in strategy). These results support out original hypothesis of a change in the basic scan pattern as mental workload is increased, but indicates the effect is much more evident in pilots of moderate skill.

DISCUSSION

Our results suggest that in a skilled task such as piloting, in which visual scanning plays an important role, the scanning behavior may serve as an indicator of both workload and skill. Before discussing some of the implications of this finding, it is important to note from the outset that the results presented do not seem to support the notion of an accurate, absolute measure of workload. However, a quantitative relative comparison of mental workload under varied conditions does appear to be feasible.

One implication of the results applies to the estimation of workload of some new procedure which may have several possible levels. In many cases, test pilots with superior flying skills are utilized in the estimation or measurement of workload. This often results in equivocal results when comparing alternative procedures, controls, or displays. The present results suggest that different levels of workload may be difficult to measure in such subjects since they appear to be less sensitive to increased difficulty. Our results suggest that pilots of moderate skill are more sensitive to the verbal mental loading task. Thus, if one is concerned with the question of the effect of changing the level of difficulty of some task, then as one step in the evaluation, the use of pilots of intermediate skill at several mental loading levels would seem appropriate since their behavior (visual scanning and performance) will be altered more as a function of the mental loading task than will that of more skilled pilots.

Another possible application may be the assessment of pilot skills. The results have suggested that there is a relationship between the scanning behavior of the pilot and his skill level. The obvious place one might use this technique is in training. One may hypothesize that, as a pilot's skills develop, his visual scanning behavior will be less and less affected by non-visual increments in workload. Specifically, it appears that as skill increases, the percentage of long dwells decreases for a particular mental loading level. The scan pattern used during a fixed maneuver is also unaffected by mental loading at higher skill levels. This finding might be utilized in assessing pilots' currency, competency, and level of skill; the technique might be used to pinpoint areas which may require additional training or practice.

The work described here has barely scratched the surface on the issue of prediction of performance via workload measurement. The results suggest that this will be possible in at least some types of situations. In order to examine this matter carefully, performance on both the piloting task and the verbal mental loading task must be closely monitored.

SUMMARY

This paper describes some of the results from an experiment designed to examine the relationship between pilot visual scan of instruments and mental workload. It was found that a verbal loading task of varying difficulty causes pilots to stare at the primary instrument as the difficulty increases and to shed looks at instruments of less importance. The verbal loading task also affected the rank ordering of the scanning sequences. By examining the behavior of pilots of widely varying skill levels, it was suggested that these effects occur most strongly at lower skill levels and are less apparent at high skill levels. A graphical interpretation of the hypothetical relationship between performance, workload, and skill is introduced and results form the preliminary experiments are presented to support this interpretation.

REFERENCES

1. Bergeron, H. P. "Single Pilot IFR Autopilot Complexity/Benefit Tradeoff Study". AIAA Aircraft Systems and Technology Meeting, Anaheim, California, August, 1980, AIAA Paper #80-1869.

2. Harris, R. L., Sr. "What do Pilots See in Displays?", presented at the Human Factors Society Meeting, Los Angeles, California, October, 1980.

3. Spady, A. A., Jr. "Airline Pilot Scan Patterns During Simulated ILS Approaches", NASA Technical Paper 1250, October, 1978.

4. Wittenborn, J. R. "Factorial Equations for Tests of Attention", Psychometrika, 8(1):19-35, March, 1943.

TANKER AVIONICS AND AIRCREW COMPLEMENT EVALUATION

by

Richard W. Moss
Air Force Wright Aeronautical Laboratories (AFWAL/FIGR)
Wright-Patterson Air Force Base
Dayton, Ohio, 45433, USA

and

Gregory J. Barbato
The Bunker Ramo Corp.
4130 Linden Avenue
Dayton, Ohio, 45432

Summary

This paper describes an effort to determine control and display criteria for operating SAC's KC-135 tanker with a reduced crew complement. The Tanker Avionics and Aircrew Complement Evaluation (TAACE) Program was a four phase effort addressing the control and display design issues associated with operating the tanker without the navigator crew position. Discussed are: the mission analysis phase during which the tanker's operational responsibilities were defined and documented; the design phase during which alternative crew station design concepts were developed; the mockup evaluation phase which accomplished initial SAC crewmember assessment of cockpit designs; and the simulation phase which validated the useability of the crew system redesign. The paper also describes a recommended crew station configuration and discusses some of the philosophy underlying the selection of cockpit hardware and systems.

PREFACE

Recognizing the potential for significant cost savings to the Air Force if the navigator position could be eliminated from the KC-135 tanker flight crew, the United States Air Force directed that a program be undertaken to determine the feasibility of replacing the navigator with avionics and/or other cockpit modifications (Ref. 1). In the Spring of 1978, the Air Force Wright Aeronautical Laboratories, Flight Dynamics Laboratory, started a two and one-half year effort aimed specifically at the crew system design issues associated with removing the navigator position from the crew but retaining the pilot, copilot, and boom operator. The work, known as the Tanker Avionics and Aircrew Complement Evaluation (TAACE) Program, was a four phase effort commencing with an analysis of the tanker's mission and associated crew responsibilities. The Program then progressed to a design phase during which several alternative crew system concepts were developed. During the third phase, a series of mockup evaluations were conducted to determine the crewmember acceptance of the different ideas, while the fourth phase accomplished a simulator validation of the findings of the previous mockup activity. The Program results indicated that operation with a reduced crew complement is a viable option if suitable equipment changes are made; specific crew station concepts and control and display subsystems were identified which offer the potential for operating the tanker without the navigator.

Additional engineering development is currently underway and a final determination of the feasibility of operating the tanker without a navigator has yet to be made. The following sections of this paper discuss the activities involved in accomplishing each of the four program phases, concluding with a brief philosophical overview of the approach used in performing the program.

1. PHASE I - MISSION ANALYSIS

The first ten months of the TAACE Program were devoted to documenting SAC's tanker mission and identifying the duties and responsibilities of each of the four flight crewmembers. The intent was to establish as complete a description of the entire system as possible; the vehicle, its subsystems, the mission tasks, and the operational context in which the system was employed. Heavy reliance for collecting this data was placed on in-flight observation of training and operational flights and crewmember interviews. A total of 25 in-flight observations were made, ranging from Combat Crew Training flights to European theater refueling exercises. During these flights, crewmembers were given opinion questionnaires to fill out aimed at obtaining subjective workload estimates as a function of both crew position and mission segment. Figure 1 summarizes the workload ratings for all crewmembers for all observation flights and permits a subjective comparison of the perceived levels of workload for selected segments of the tanker's mission.

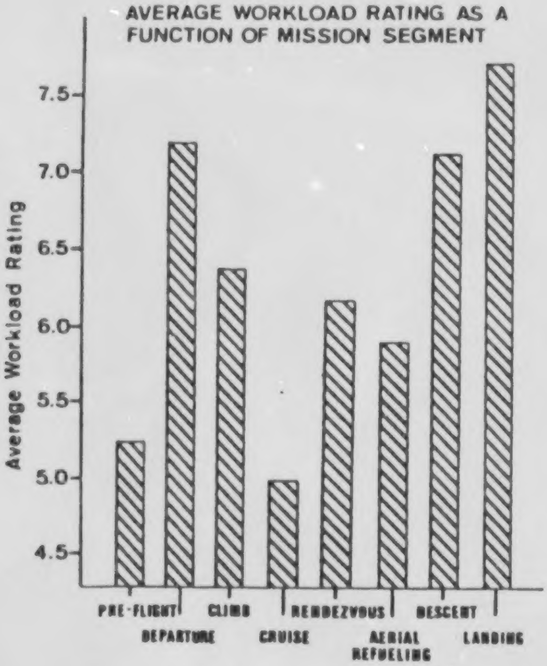


Figure 1. Average Workload Rating as a Function of Mission Segment.

In addition to this subjective workload assessment, a great deal of descriptive information was obtained during each flight identifying the activities engaged in by the crew during the different mission segments. This was done so that a complete statement of tanker operations and crewmember responsibilities was available to assure that necessary capabilities were not inadvertently eliminated from the new, three crewmember configuration. The descriptive information was collated and used to invent a hypothetical mission detailing an operational employment of the tanker. This document, known as the mission scenario, was drafted at Wright-Patterson AFB, and then carefully coordinated with Headquarters SAC personnel and operational crews to assure that it was realistic and conformed to the Command's views and needs. After several iterations, a version was established that was representative of the types of events encountered by the KC-135 during day-to-day operations. Included in the scenario for completeness were selected system failures, various weather conditions, and different levels of crew workload. A great deal of care was exercised in creating the scenario as it established the standard against which any new designs would be measured. It was realized that unless an accurate test of the new concept was available, little assurance as to the useability of the new system could be established.

The scenario was documented in three ways; by altitude profiles (Figure 2, 3, and 4) a time line, a portion of which is shown in Figure 5, and a narrative.

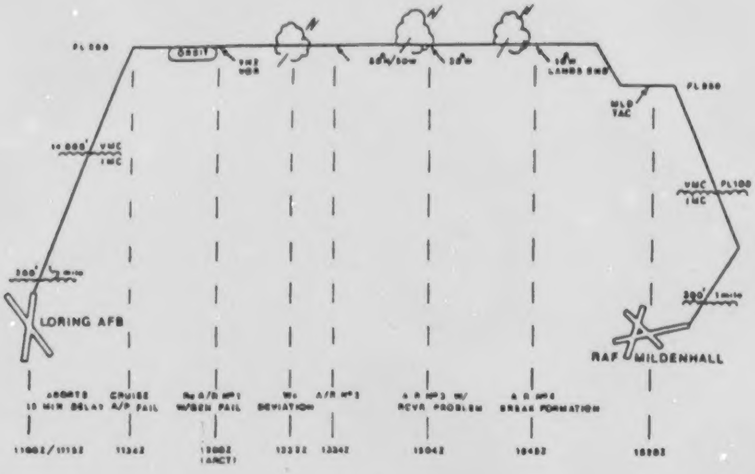


Figure 2. Profile of First Flight - Loring AFB to Mildenhall RAFB.

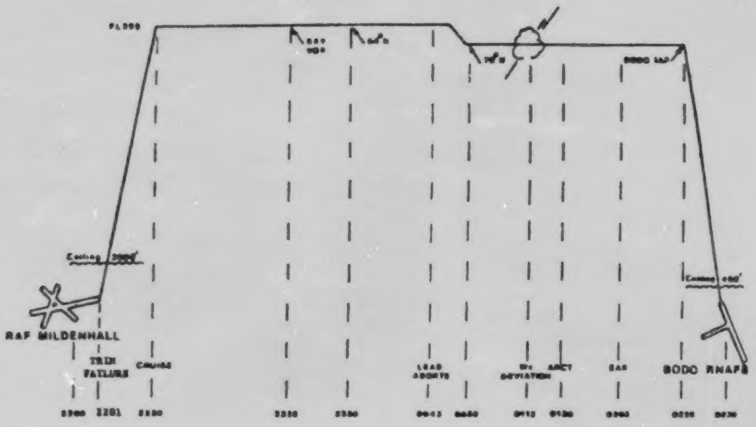

RAF MILDENHALL
TRIM FAILURE
CRUISE
LEAD ABORTS
DEVIATION
ARCT
BODO RNAFB
2201

Figure 3. Profile of Second Flight - Mildenhall RAFB to Bodo RNAFB.

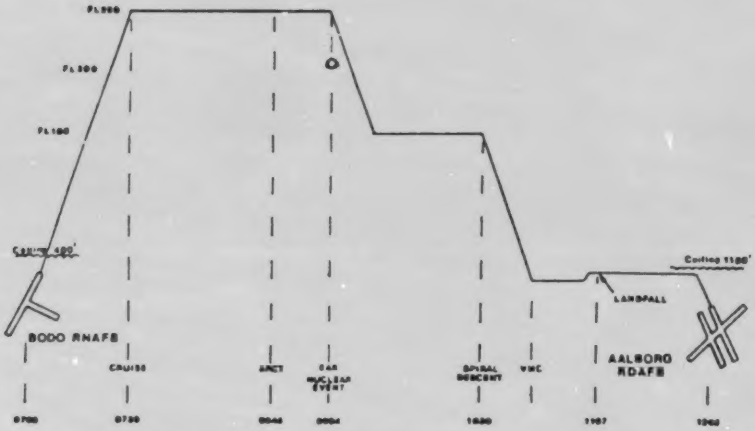

BODO RNAFB
CRUISE
ARCT
NUCLEAR EVENT
SPIRAL DESCENT
VMC
LANDFALL
AALBORG RDAFB

Figure 4. Profile of Third Flight - Bodo RNAFB to Aalborg RDAFB.

ELAPSED TIME	GMT TIME	PILOT	COPILOT	BOOM OPERATOR
02 06	1146	MONITORS AUTO PILOT HEADING AND ALTITUDE CONTROL CONFIRMS SYDNEY ETA at 1205	ACKNOWLEDGES MONCTON REQUEST FOR IFF IDENT AND TRAFFIC INFORMATION	DECODES HF MESSAGE
02 11	1151	MONITORS AUTO PILOT AND SLIGHT COURSE CORRECTION CHECKS WIND AND DRIFT	ACKNOWLEDGES ESSO 25 A/R PROGRESS	MAKES COMM LOG ENTRY
02 24	1204	MONITORS MAP DISPLAY POSITION NOTES FUEL STATE	NOTIFIES ESSO 25 FOR TURN PREPARATION OVER SYDNEY	MAINTAINS OUTSIDE VISUAL WATCH
02 25	1205	NOTES SYDNEY STATION PASSAGE TURNS ON COURSE WITH AUTO PILOT	NOTIFIES ESSO 25 FOR SYDNEY PASSAGE AND TURN MANEUVER ACKNOWLEDGES ESSO 25 A R PROGRESS	UPDATES POSITION ON IN MAP MAKES ENTRY ON FUEL AND COMM LOG
02 26	1206	MONITORS AUTO PILOT ETA TO ST JOHNS GS TAS WINDS DRIFT AND PRESENT POSITION	ACKNOWLEDGES MONCTON TRAFFIC AND IFF CODE CHANGE MONITORS FUEL STATE MAKES ENTRY INTO NAV LOG	MAINTAINS OUTSIDE VISUAL SCAN

Figure 5. Portion of time-line identifying crew tasks during cruise segment.

Briefly, the scenario describes the events and tasks accomplished by a single tanker during three operational flights over a period of three days. The first flight is a fighter deployment to Europe supported by a five ship tanker force. The described tanker is briefed to be the number two aircraft in the tanker formation, but due to system malfunctions experienced during the take-off roll, must delay on the ground for minor repair. The resulting late departure generates out-of-the-ordinary navigation tasks, created in the scenario as the first major test of the system's ability to cope with unexpected situations. Eventually, the aircraft joins with the previously departed formation and proceeds to Europe. During the flight, various tanker system failures occur, as well as encounters with adverse weather and receiver refueling malfunctions. An uneventful IFR recovery in England terminates the first flight. The second flight is an Emergency War Order (EWO) mission, in which the described tanker is number two in a two ship formation launched from Mildenhall RAFB to join with two B-52 bombers over the North Atlantic. During the cruise segment to the rendezvous, the lead tanker experiences unresolvable system failures requiring single ship completion of the remainder of the mission. Once again, tanker malfunctions dictate additional workload, including tanker control of the rendezvous with the receivers and minimum fuel landing at the recovery base. The third and final flight is also a two ship mission, launched from Bodo AFB, Norway, to support tactical air operations in Eastern Europe. An extremely high workload situation is established during the orbit and subsequent refueling operations with the mission unexpectedly terminated by the detonation of a nuclear device somewhere in the vicinity of the aircraft. A complete description of the mission, including the entire narrative and time line, is provided in the TAACE Program Final Report (2). Figure 6 is an artist's overview of the entire scenario.

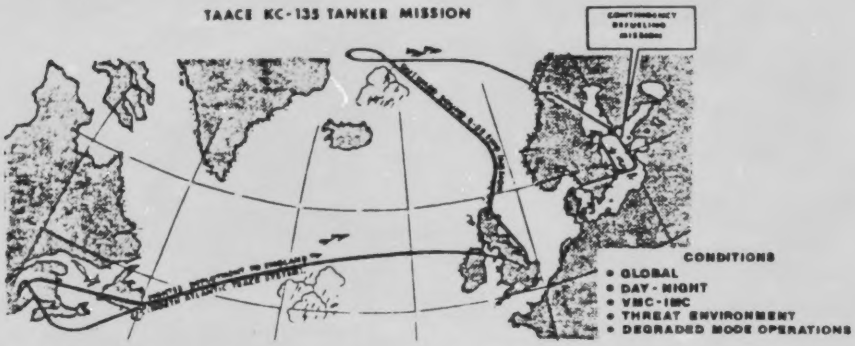

TAACE KC-135 TANKER MISSION
CONTINGENCY REFUELING MISSION
CONDITIONS
• GLOBAL
• DAY - NIGHT
• VMC - IMC
• THREAT ENVIRONMENT
• DEGRADED MODE OPERATIONS

Figure 6. Overview of entire mission scenario.

2. PHASE II - CREW SYSTEM DESIGN

After thorough analysis of the scenario, a tentative list of crew station revisions was made identifying the elements of the existing system that might be altered in order to affect the elimination of the navigator from the crew. In addition, readily available systems and control and display components that would have to be added to the vehicle were also catalogued and their suitability assessed. This hardware oriented activity provided the transition to the second phase of the program, the development of specific design concepts for flying the tanker without a navigator.

The second phase of the TAACE Program dealt with the development of crew system concepts that would provide for operation of the vehicle, monitoring of subsystem's performance, accomplishment of the refueling tasks, and complete and safe accomplishment of the navigation job without the benefit of the navigator crewmember. It was clear from the outset that particular attention would have to be paid to the reworking of the navigation subsystem. However, since the level of automation and integration needed to eliminate the navigator from the crew was not known beforehand, three alternative crew station designs were developed. They differed from each other in several ways: the degree of change or update to the existing tanker; the amount of automation incorporated in the design; and the costs associated with retrofitting the fleet. These three designs were named the MINIMUM, MODERATE, and MAJOR updates.

The MINIMUM update (Figure 7) redesign was developed as the least expensive alternative to operating the tanker without a navigator.

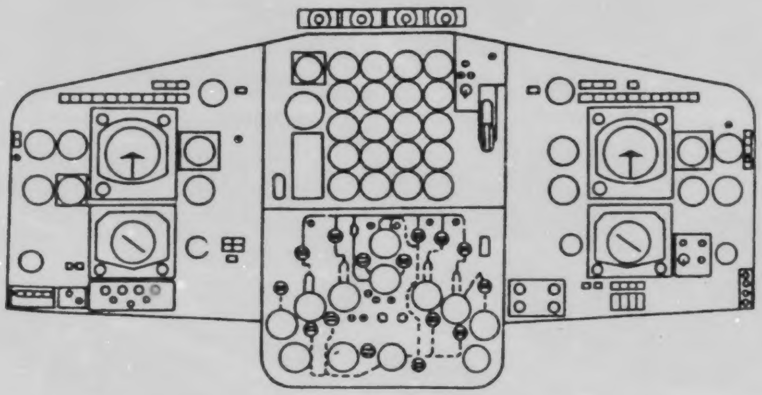


Figure 7. MINIMUM Update Redesign.

It incorporated very minor changes to the baseline configuration and is characterized more by the rearrangement of on-board controls and displays than by the addition of new systems. It was felt that the minimum update represented the least sophisticated (and least costly) concept with any chance at all of providing the necessary functions and capabilities compatible with a reduction in crew size. In this sense, it was one end of a sophistication/cost/capability continuum by which the three designs could be compared.

The MODERATE update (Figure 8) was developed as the most logical candidate, one having a moderate blend of the new systems, rearranged existing systems, and modified crew station geometry. This design - characterized by the addition of several major cockpit subsystems including integrated caution

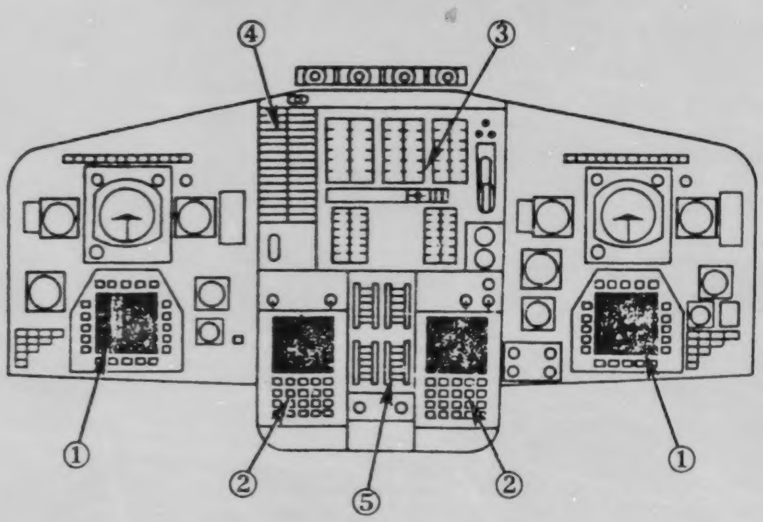


Figure 8. MODERATE Update Redesign showing (1) Pilot and Copilot Electronic HSI's, (2) Navigation Management Subsystem CRT/Keyboard Units, (3) Reconfigured Engine Instrument Panel, (4) New Caution/Warning Panel and (5) Fuel Quantity Instruments.

and warning annunciator lights, vertical scale engine instruments, electronic displays to replace the electro-mechanical HSI's, and two electronic display/keyboard units (Figure 9) through which the crew interacted with the new navigation management of existing controls and display - was felt to provide the necessary level of crew support to assure successful mission completion.

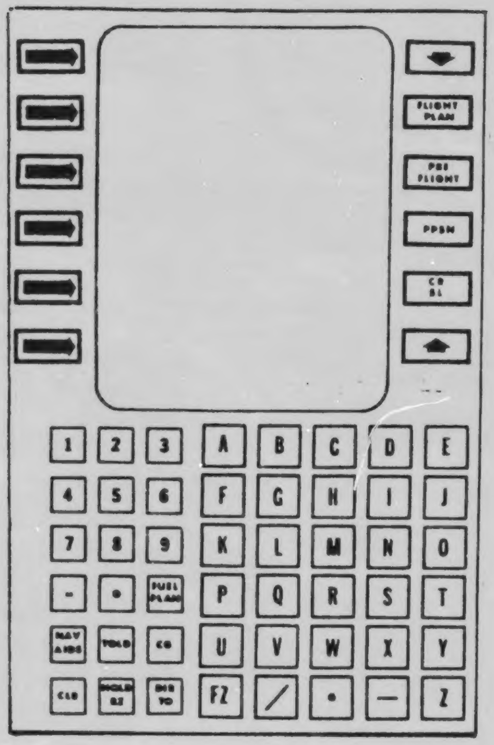

FLIGHT PLAN
PRE FLIGHT
PPSN
1
2
3
A
B
C
D
E
4
5
6
F
G
H
I
J
7
8
9
K
L
M
N
O
-
FUEL PLAN
P
Q
R
S
T
NAV AIDS
TOLD
U
V
W
X
Y
CLR
DIR TO
FZ
/
•
—
Z

Figure 9. Navigation Management CRT/Keyboard Unit.

The third and final design, the MAJOR update (Figure 10) was developed to consider a configuration felt to be more than adequate to do the job.

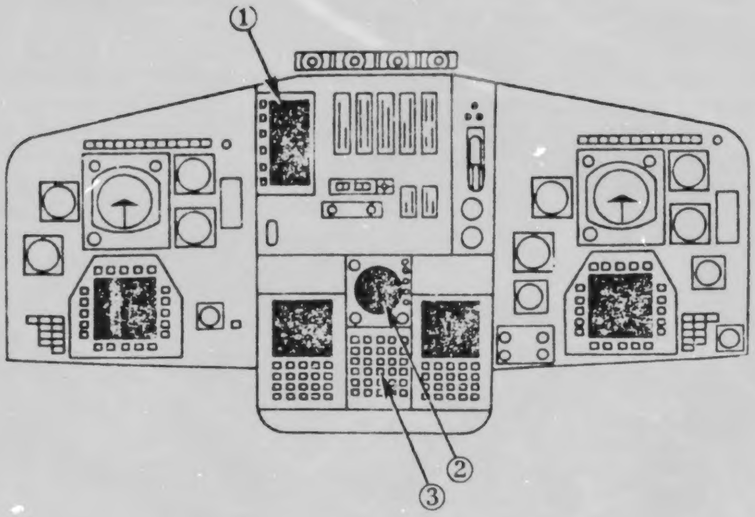


Figure 10. MAJOR Update Redesign showing (1) Automated Checklist Readout, (2) Dedicated Weather Radar Scope and (3) Fuel System Switch Matrix.

From the outset, it was assumed that the MAJOR update was more sophisticated than was absolutely necessary. In addition to the modifications advocated for the MODERATE update, the MAJOR update included a completely automated fuel management system, integrated tuning of all voice radios, and computer storage and electronic display of checklists and emergency procedures. In a sense, the MAJOR update represented the opposite end of the continuum from the MINIMUM update, the two bounding the design problem, thereby placing limits on the extremes of control and display sophistication to be explored. In the middle, representing what appeared to be a logical trade-off between the austerity of the MINIMUM update and the extravagance of the MAJOR update, was the MODERATE update.

The design phase of the TAACE Program was not confined to the development of the three designs on paper. As part of the process of creating the different configurations, a full-scale mockup facility (Figure 11) was constructed which was used to verify the suitability of control and display location, crew station egress and ingress, and overall crew coordination. The mockup was an accurate replica of the KC-135 cockpit and played a fundamental role in the design process. As the three configurations were evolving, each in its turn was laid out in the mockup and assessed for adequacy from the designers point

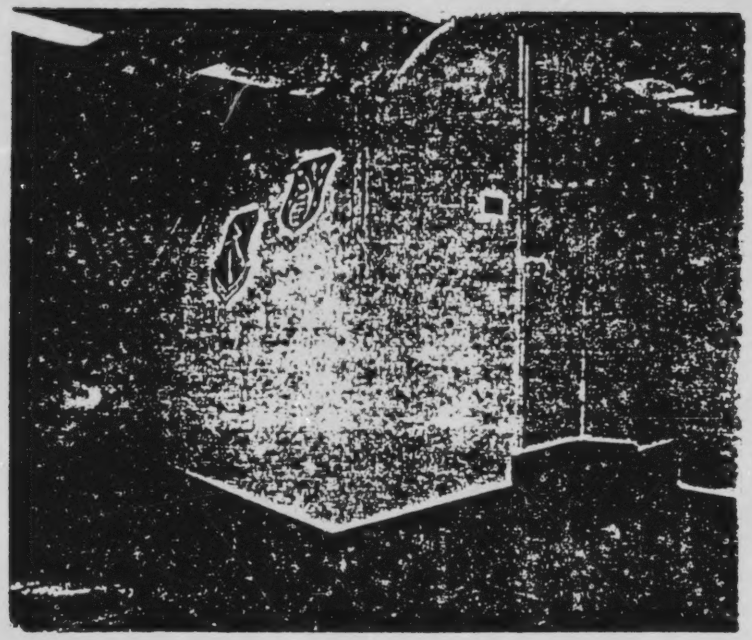


Figure 11. Exterior of full-scale mockup used during TAACE program.

of view. Thus, during the course of the design phase, each configuration was continually examined and refined as appropriate. After the configurations were fairly well established, cockpit checklists and procedures were developed, an integral part of the design task, directed toward establishing crew roles/duties and efficient crew system functioning. Checklists and procedures are the operating instructions that inform the crew of the techniques for using the system during both normal procedures and emergencies, and operating procedures often represent a direct link between the design community and the user. During the TAACE Program, cockpit procedures were extremely important; SAC has established very specific ways of performing its mission and the already-in-existence operating procedures for the tanker in the context of the refueling mission often established as definitive a set of design requirements as did the crew size issue itself.

When the point of diminishing returns was reached-the pilot where interations to the designs produced very little real change-the program progressed to the third phase, the mockup evaluation of the three configurations.

3. PHASE III - MOCKUP EVALUATION

The third phase of the TAACE Program performed an initial set of evaluations of the three designs; an initial assessment of the utility and pilot acceptability of each of the three configurations from the operator's point of view. The mockup evaluation returns the focus of activity to the user, providing another opportunity for user inputs to the development process. The intent is to create a structured situation during which operationally experienced crewmembers are given the chance to assess how well the design can support safe and effective accomplishment of the mission.

A total of nine fully-qualified, operationally experienced SAC tanker crews participated in the mockup evaluation. The general sequence of activities for each crew included attending ground school, receiving realistic mission briefings. Three crews participated in the evaluation process at one time. Starting on a Monday morning, and finishing on a Friday evening, it typically took five days to complete all scheduled activities. Thus, a total of three weeks were needed to collect all the mockup evaluation data. On the first day of each week, three crews were introduced to the TAACE Program at which time the goals and objectives of the work were explained. This rather informal session helped establish rapport with the user and generate enthusiasm for the process wherever possible. It was pointed out that the mockup activities to follow were a significant departure from the traditional cockpit design process, and that the significance was due primarily to the user's involvement and willingness to "role play" during upcoming mockup flying sessions.

Following the orientation briefing, the crews attended several hours of ground school, classroom type sessions during which the cockpit designs were discussed and explained. A great deal of time was spent briefing the new systems and their operating procedures; line crews are often not familiar with the new concepts or hardware and an effective evaluation demands that the crewmembers thoroughly understand how the equipment operates. Ground school is a relatively formal give and take of information with the crews learning the new configurations and how the equipment is intended to help them perform the mission, while passing along insight into good and bad features of the different ideas. Often, unanticipated uses for features of the design are discovered while deficiencies or illogical procedures may also be noticed. Frequent use is made of the mockup to familiarize the crewmembers with the location of equipment; during the flying sessions they are asked to move their hands from control to control in order to assess reach envelopes and accessibility, and awareness of the location of equipment facilitates this assessment.

Actual role playing in the mockup took place after the ground school sessions were completed. As was mentioned earlier, nine SAC crews participated in the mockup evaluation phase of the program. However, for the activities associated with the actual mockup role playing, each crew was treated individually. Mission briefings began the process with each crew being given flight, weather, communications, formation and refueling information in a manner similar to standard SAC tanker operational briefings. The intent was to create a mind set within the crewmembers such that once inside the mockup, they were thinking along mission lines and evaluating the designs in the context for which they were intended. Each crew received three different mission briefings, one for each of the three flights developed during the mission analysis phase of the program. During the briefings, the crews were encouraged to ask questions about the mission and to make sure they understand the job to be done. They were informed that they were expected to perform all communications tasks that would ordinarily take place during the mission, maintain an awareness of their navigation situation, calculate all parameters e.g., take-off and landing data, estimated times-of-arrival, etc., and generally perform all crew station duties demanded of the mission or the systems in the crew station. Since some emergencies were to occur during each flight, the crews were also reminded of their responsibility to be familiar with all checklists and emergency procedures.

After each mission briefing the crew entered the mockup, climbed into the seats, and started the role playing exercise. They donned headsets and communicated over a functional intercommunications system, beginning the process with the appropriate checklist for the mission being flown. Outside the mockup, there was an experimenter's console manned by two test engineers responsible for monitoring the mockup session and providing external voice communications in response to either mission situations or crew initiated commands. Typically, the test engineers assumed the voice communications for such external agencies or personnel as ground control, tower, departure control, Air Traffic Control, SAC Command Post, fire guard on the ground, and other aircraft in the formation.

The test engineers also paced the mockup flight by providing indications of mission progress in the absence of functional flight or navigation instrumentation. For example, after the crew had completed the "Line-Up" checklist and simulated a take-off, a test engineer, speaking as Departure control, would call the aircraft and say that radar contact had been established. This sort of feedback helped keep the mission moving and the crew oriented, and was provided throughout the role playing exercise. Finally, the test engineers moved the flight ahead of its normal time line in order to eliminate long uneventful segments of the scenario where, because of the lack of operable equipment, the crewmembers would not be learning anything new about how the system was intended to operate. To effect a "move ahead", a test engineer would inform the crew that they had just completed a specified series of tasks and that the flight was being moved ahead to test the useability of the system under another set of conditions. The crew would then be briefed on the new conditions; e.g., altitude, heading, airspeed, GMT, weather conditions, current agencies in communications contact, and upcoming events, such as passing a specific radio aid to navigation. During the first flight, for example, the crew was permitted to accomplish several cockpit checklists, prepare for and execute the take-off, and join-up and then establish themselves mentally at cruise altitude before the first "move ahead" took place. The flight was then advanced to a position over the North Atlantic, just prior to one of the Mid Atlantic air refuelings. Approximately 30 minutes of air refueling activity was performed after which the flight was moved ahead again, this time to a point approximately 30 minutes prior to touchdown at Mildenhall RAFB. The crew was permitted to complete the final segment in real time, making radio contact with all the necessary controlling agencies in England, performing necessary descent and before landing checklists, and simulating the approach and touchdown.

The overall role playing exercises encompassed a total of twenty-seven flights. Every crew flew each of the three mission flights once, using one of the three designs; thus, each crew flew all tasks and used all designs, but did not experience all the possible combinations of design coupled with flights. Over the course of the total mockup phase however, each of the nine possible combinations occurred three times.

Immediately after the flying sessions, the crews were handed questionnaires asking their opinion about the useability of the design just flown to do the job just completed. The intent was to not only determine which of the three designs was best suited to a removal of the navigator, but how, if at all, their suitability varied as a function of mission tasks.

The mockup flying and subsequent evaluations provided by the SAC crews identified both good and bad features of all three designs, resulting in the development of a fourth configuration, a composite of elements from each of the original three. The detailed results of the Mockup phase are documented in an Air Force Wright Aeronautical Laboratories, Technical Report, AFWAL-TR-80-3030 (Vol. 1) (Ref. 2). A summary of those results is provided here.

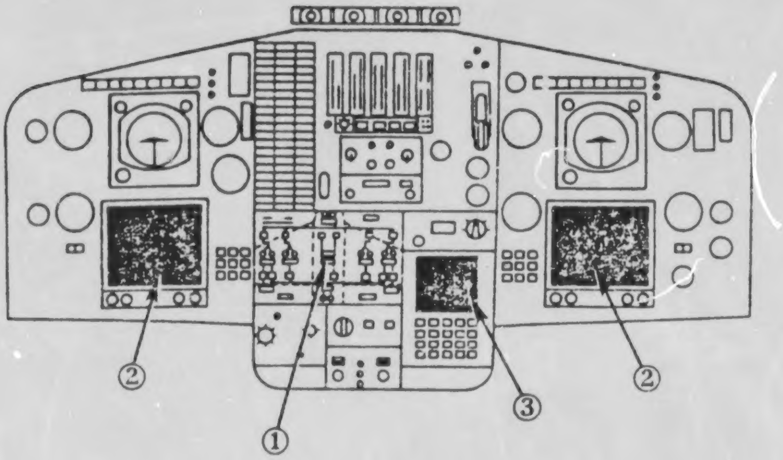


Figure 12. COMPOSITE Redesign. Note (1) Schematic representation of fuel system, (2) Retained pilot and copilot electronic HSI's, and (3) Navigation Management CRT/Keyboard Unit. Second unit is installed on an isle stand, aft of throttles, not shown.

The desired COMPOSITE design (Figure 12) was characterized by a navigation management system similar to the one defined in the MODERATE update, capable of displaying at least six upcoming waypoints at a time, computing take-off and landing data, storing holding or rendezvous patterns for air refueling operations and generating flight plan data such as time-to-go, or distance-to-go. None of the fuel management concepts included in the original three designs were judged suitable by the crews. Instead, a reduced version of the existing tanker fuel panel, with essentially the same capabilities as the existing panel, is included in the COMPOSITE. The most significant feature of the composite design is the electronic display replacement for the electro-mechanical HSI. This device displays, at pilot discretion, conventional horizontal situation data (Figure 13), computer generated map data (Figure 14), weather radar information (Figure 15), holding or rendezvous patterns (Figure 16 & 17), and other information typically provided by the navigator. The electronic HSI's, coupled with the navigation management subsystem,comprise the heart of the avionics needed to operate the tanker without the navigator crew position. The management system computes navigation data, manages some of the navigation radios, and generally performs the calculations accomplished by the navigator, while the electronic displays present to the pilot and copilot the information generated by the computer.

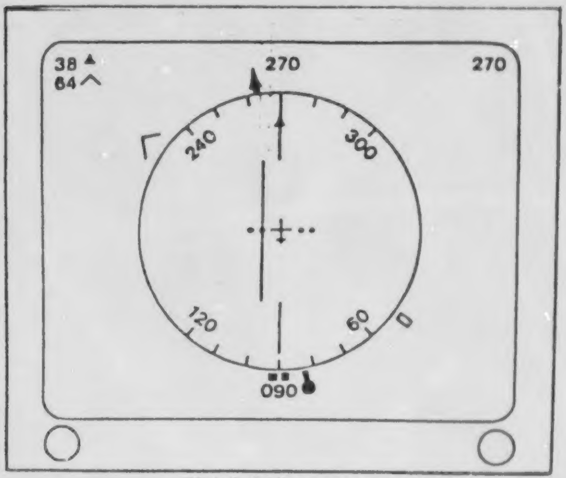


Figure 13. Electronic HSI, showing conventional horizontal situation information.

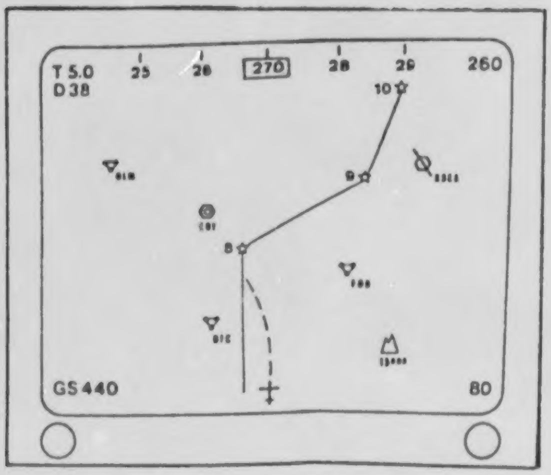

T 5.0
D38
25
26
270
28
29
260
10
9
8
GS440
80

Figure 14. Electronic HSI, showing map format. Parameters shown clockwise from upper left-hand corner, are Time (T) and Distance (D) to next waypoint, current aircraft heading (270), course between last waypoint and "To" waypoint (260) map range in nautical miles (80) and current ground speed (GS).

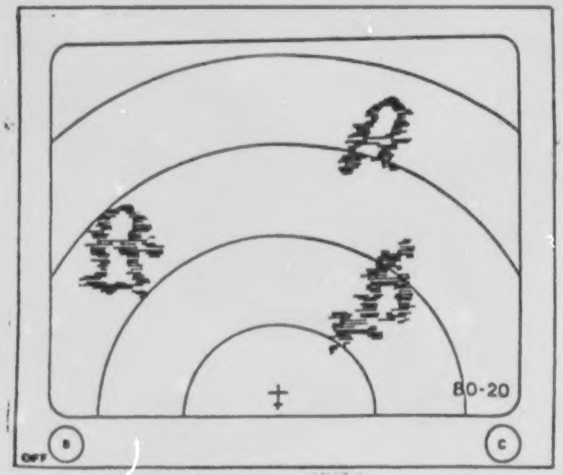


Figure 15. Electronic HSI, showing, weather radar information.

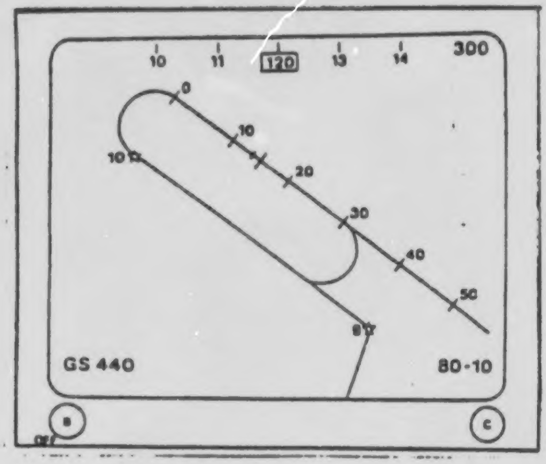

10
11
120
13
14
300
0
10
20
30
40
50
10
8
GS 440
80-10
B
C
OFF

Figure 16. Electronic HSI, showing holding pattern information. ORBIT shown is computer generated, based on information supplied by the crew.

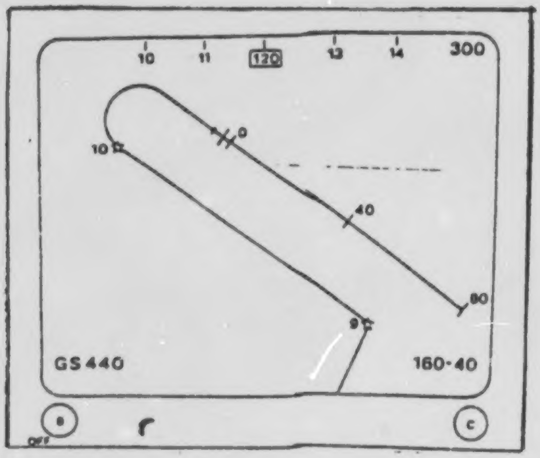

10
11
120
13
14
300
0
10
40
80
9
GS 440
160-40
B
C
OFF

Figure 17. Electronic HSI, showing air refueling rendezvous information. Note airplane symbol and 40 and 80 range ticks. Range ticks represent distance from airplane symbol, and are used to anticipate tanker 180° turn to lead receiver along refueling course.

At this time in the program, a candidate avionics package and crew station configuration had been developed, based on the using commands evaluation of several alternative crew system concepts. A paring down process was taking place, having started during the design phase, where obviously unacceptable ideas had been eliminated, and continuing through the mockup exercise where additional assessment took place. The last phase of the TAACE Program accomplished the final paring down of ideas - fine-tuning the COMPOSITE design.

4. PHASE IV - SIMULATION VALIDATION

In general, the simulation work performed to validate the crewmember acceptability of the COMPOSITE design followed the same procedures employed during the mockup activity; SAC crewmembers were obtained to "fly " the system under simulated conditions, questionnaire data was obtained to record crewmember opinion, and a final set of recommended crew system control and display design criteria were generated.

The simulation facility used for the TAACE Program consisted of a cockpit cab (Figure 18), a visual projection system, a test engineer's console, a simulated boom operator's console, and supporting computer systems. The simulator cab was a geometric duplicate of the KC-135 cockpit outfitted with operable flight instruments, flight controls, lighting, communications radios, and navigation management subsystem (Figure 19).

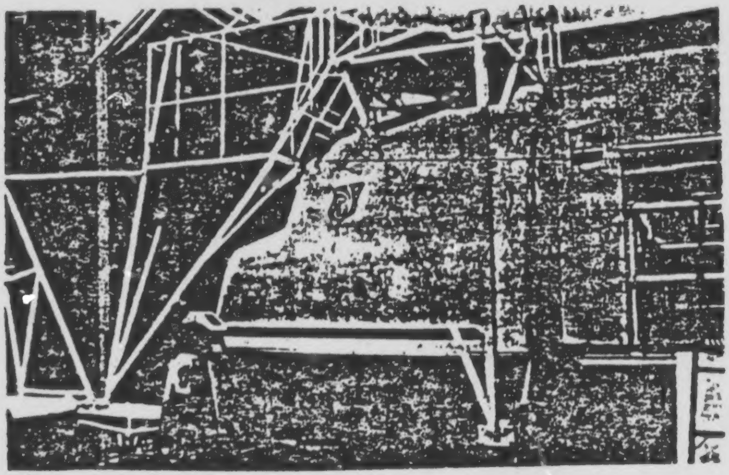


Figure 18. Exterior of simulator cab used during TAACE Program.

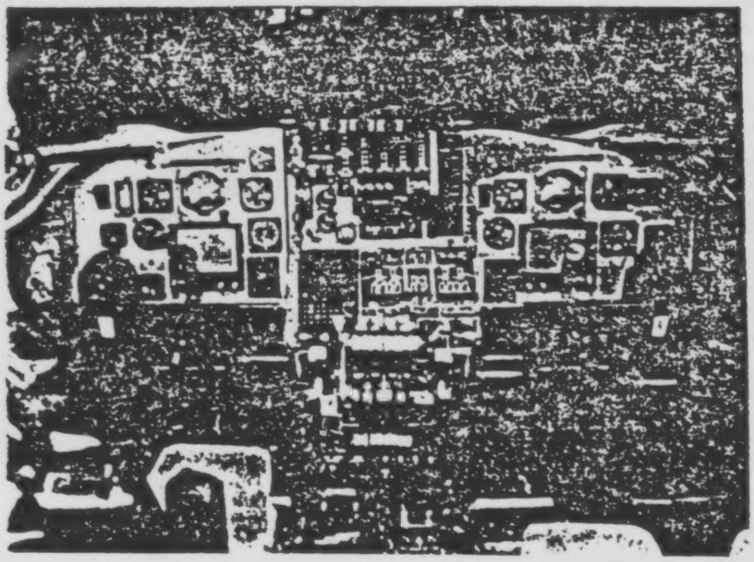


Figure 19. Interior of simulator cab configured with "COMPOSITE" design.

The crew could program any flight plan into the navigation system and through proper switch selection have relevant navigation data appear on the electronic HSI and flight director. The crew could then fly the flight plan and have computed for them pertinent navigation information. The system also displayed ground speed, true airspeed, time and distance to go to future waypoints, holding and rendezvous patterns, and other data deemed necessary as a function of the mockup evaluation work. To enhance the realism of the simulation, environmental conditions of day, night, and engine sound were provided along with a visual presentation of airport features for take-off and departure operations. The test engineer's console fulfilled the same functions as did the console used during the mockup work, with the addition of cockpit repeater displays to show information selected by the crew. Finally, a boom operator's console was fabricated to provide the boom operator, who participated as a member of the crew, with tasks to perform which simulated his activities during the air refueling segments of the flight.

The simulator sessions, flown by four crews not previously involved in the program, were conducted in the same way as the mockup sessions with each crew receiving mission briefings, entering the cab, flying the mission, and completing questionnaires. Since the real world flying task was simulated, the pilots, in addition to their mission responsibilities for navigation and subsystem monitoring, also had to fly the airplane. Thus, the single biggest difference between the mockup and simulation phases was the increased workload levied upon the crew, and hence, a more valid assessment of the design's useability. Another difference was the availability of objective system performance data. Prior to the simulator flights, the optimum mission profile consisting of headings, altitudes, arrival times, etc., was programmed into the computer for comparison with the crew generated values for the same parameters. This data was then analyzed to augment the pilot questionnaire data in making decisions about the suitability of the avionics systems flown by the crew.

In contrast to the large differences between the MINIMUM, MODERATE, and MAJOR crew stations that were evaluated during the mockup work, resulting in the COMPOSITE design, the simulation phase attempted to assess the relative goodness of much more subtle crew system alternatives. For example, after the mockup work, it was unclear as to exactly how much flexibility the crew should have in selecting for display raw navigation sensor data. The mockup evaluation was simply not sensitive enough to determine this. Consequently, during the simulator flights the crews, although always using the same equipment, were given different levels of capability with which operate. During one leg of a flight, for example, they might fly with only one navigation management system electronic display/keyboard unit operable, while at other times, have access to two units. Thus, the simulation phase was a continuation of the paring down process started during the design phase, with the level of emphasis placed upon finer details of equipment operational capability or arrangement.

The alterations in the COMPOSITE design needed to make it acceptable to the SAC crewmembers who evaluated it were not nearly as extensive as those made to the MINIMUM, MODERATE, and MAJOR designs to generate the COMPOSITE configuration. As expected, changes dealt primarily with details of display format, information content, or of the navigation subsystem. For example, it was felt that the Bearing Distance Heading Indicator should have the capability to display either TACAN bearing or flight plan waypoint bearing, as selected by the pilots; that when radar imagery was displayed on the electronic HSI certain navigation parameters be retained in the corners of the display; and that waypoints be retained in the flight plan until erased by the crew rather than eliminated automatically as used.

5. CONCLUDING REMARKS

The TAACE Program is significant because it demonstrates an objective, structured process for dealing with the very complex issues of crew system design and evaluation. Figure 20 overviews this process starting with the mission analysis, progressing to the preliminary design phase, and moving through the various levels of user-in-the-loop evaluation.

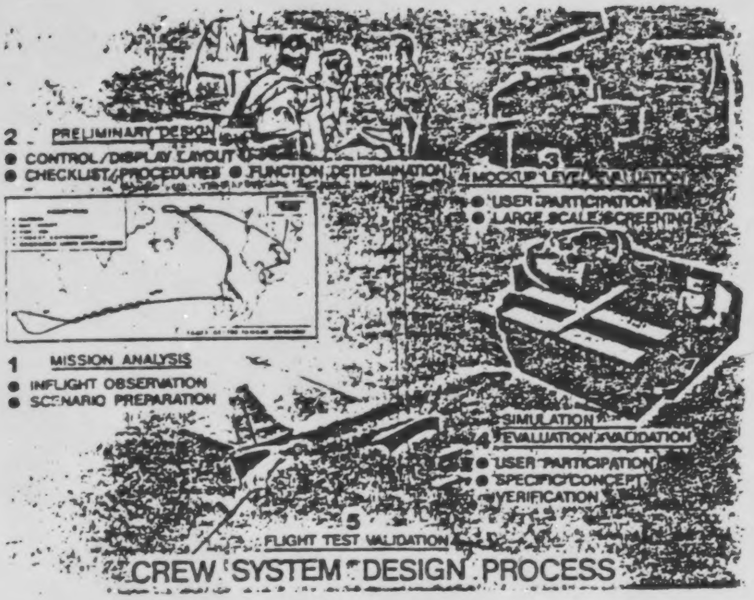

2 PRELIMINARY DESIGN
● CONTROL/DISPLAY LAYOUT
● CHECKLIST/PROCEDURES ● FUNCTION DETERMINATION
3 MOCKUP LEVEL EVALUATION
● USER PARTICIPATION
● LARGE SCALE SCREENING
1 MISSION ANALYSIS
● INFLIGHT OBSERVATION
● SCENARIO PREPARATION
SIMULATION
4 EVALUATION VALIDATION
● USER PARTICIPATION
● SPECIFIC CONCEPT VERIFICATION
5 FLIGHT TEST VALIDATION
CREW SYSTEM DESIGN PROCESS

Figure 20. Crew system design process, showing major activities pursued during design and evaluation of new cockpit.

It is important to note the sustained reliance upon the user throughout the entire methodology, from the intense interface during the development of the design scenario to the mockup, simulator, or flight test work. It is also important to note the continued use of a validated, comprehensive scenario. At the outset, the scenario represents a definition of the problem being solved by the evolving crew system. Later in the process, it represents the criteria against which the design's useability is measured. Finally, the method provides for the collection of subjective pilot opinion data, objective system performance data, and other factors that can all be brought to bear on the decision making process. This was the approach used in the TAACE Program and it has proved compatible with the other on-going processes currently in use within the United States Air Force for developing and procuring new weapon systems.

REFERENCES

1. KC/C-135 Avionics Modernization Program Management Plan. Strategic Systems Program Office, 24 April 1978, Aeronautical Systems Division, Wright-Patterson AFB, Ohio

2. Barbato, Gregory J., Madero, Ralph P., and Sexton, George A., Tanker Avionics/Aircrew Complement Evaluation (TAACE) Phase 0-Analysis and Mockup Final Report - Vol. I, II, III, AFWAL-TR-80-3030, May 1980.

ABSTRACT

OPERATIONAL MONITORING IN MULTI-CREW TRANSPORT OPERATIONS

Capt. Harry W. Orlady (Ret.)

At least two kinds of failures are involved in a majority of the flight crew related operational anomalies which occur in air transport operations. The first is the failure in expected performance by the pilot flying which leads directly to the operational anomaly. The second is the failure of the other pilot to detect the performance deviation or deviations responible for the "unwanted occurrence" in a timely and effective manner.

This paper is concerned with the second failure. It considers operational monitoring by the pilot not flying as a sub-system of the total operational task and discusses selected aspects of monitoring from this viewpoint.

AN ORGANIZATION DEVELOPMENT APPROACH TO RESOURCE MANAGEMENT IN THE COCKPIT

Linda Orlady Rings
Department of Aviation
The Ohio State University

With the advent of computer technology in the cockpits of modern airplanes and the pressure to reduce the size of the flight deck crew in airline and general aviation operations, an examination of the resources that yet remain available is needed both to determine whether they are adequate to solve routine and emergency flight problems and to determine how to use them effectively. There is a growing concern, particularly in general aviation, that inadequate training in resource management may be the reason for a number of accidents and incidents. This concern was formally expressed in an NTSB report of the spring, 1980 crash of a DHC-6. The report cited the lack of resource management training as a contributing factor of the fatal crash.

The resources available to the pilot can be put into three broad categories: cockpit equipment, the ATC system, and other pilots in the flight crew. Pilots are highly trained in the use of the cockpit equipment and generally use it effectively. For example, if an autopilot is available and working correctly, its use allows the pilot to spread his or her workload more efficiently and to have more time for other tasks such as scanning for other aircraft, copying clearances, etc. Air traffic control, on the other hand, is an often underutilized resource. Frequently its full potential is not realized or appreciated until an emergency situation occurs. The third resource, the right seat occupant, is often overlooked. Some left seat pilots think of the right seat occupant as simply a passenger. His operational relevance is simply a weight that shifts the center of gravity of the aircraft forward a couple of inches.

The reason that resources such as ATC and the right seat occupant are not properly used may be that we do not yet know how to use them effectively in all areas of aviation. One potential approach to a solution of this problem is Organization Development (OD), a conceptual model that has been used successfully in management science. The objective of this paper is to demonstrate how application of the problem-solving skills inherent in OD can be used to better manage the resources available in the cockpit. The paper focuses on a particular segment of the aviation industry, general aviation, which encompasses all civilian flying except that of the scheduled air carriers.

OD is a systemwide concept to increase organizational effectiveness. The organization can be any relatively autonomous group; it need not include the entire firm. In aviation the organization is the left and right seat occupants for a particular flight. The objective of an OD program is to provide first a systematic diagnosis of the organization, and secondly, a strategic plan for improvement of resource management. OD should also facilitate problem solving on the job. Its framework should allow for changes in the environment.

OD is not new to the aviation industry. Crew coordination has long been a concern of the airlines and other users of multi-person crew aircraft. The term, "resource management", has been applied more recently as better describing processes necessary for safe and efficient operation of the aircraft. These processes involve OD principles. They include not only crew coordination but also leadership and commandibility skills. Several major airlines are utilizing organization development techniques to better manage cockpit resources. An examination of these techniques may help assess their potential contribution to the non-airline sector.

American Airlines calls its interpersonal skills training program "TACT", an acronym for Transactional Analysis in Customer Treatment. The program is based on the Thomas A. Harris theory of transactional analysis and was originally developed for training customer contact personnel such as ticket agents. Approximately five years ago, the TACT program was incorporated into the upgrading program for copilots transitioning to captain status. OD would suggest that such programs are of limited value because they deal with only one section of the cockpit organization, the future captains. OD requires an integrated approach including the full crew. Failure to use the other members of the flight crew may cause incorrect perceptions of, not only their own role, but also that of the other crew members leading to an ineffective utilization of this cockpit resource. This problem has appeared in other applications of the OD approach where indoctrination was given only to top management. Others in the organization typically resist the new approach primarily from a lack of understanding.

Eastern Airlines has one of the oldest resource management programs. Its Line-Oriented Flight Training program (LOFT) was first used in the late 1950's on DC-8 and Boeing 720 series aircraft. Today Eastern's LOFT program takes a complete flight crew through a four hour simulator mission, during which they are faced with mechanical malfunctions or failures under a variety of real world operating conditions. Eastern has found that the program is an effective evaluation tool, not only of their flight crews but also of their training and operation procedures. One of their training managers has stated:

> "The simulator's ability to accurately reproduce the line pilot's normal working environment, plus the instructor's briefing prior to the period, emphasizes to the crew that they are expected to perform in the simulator exactly as they would perform in the real world. This permits us to see a more accurate picture of how the crew functions in such areas as decision-making, cockpit discipline, the Captain's command presence, crew coordination, and other resource management skills" (Beach, 1979).

This statement also points to a criticism of LOFT. Some people have suggested that LOFT is more of an evaluation program than a training program. It should also be noted that Eastern, like several other airlines, is in the process of developing a LOFT program for initial training in addition to the program for recurrent training. Furthermore, the evaluation is structured so that it is not a pass/no pass evaluation and is thus used for training as well.

These and other successful programs have several principles in common. First, the developers of the programs have had the support of top management from the beginning. Secondly, they have acquired critical support by inviting both the FAA and their pilots to participate in the development of the program and simulator missions, and in a continual review of the program to improve it.

The major effort being made by the airlines to reexamine and use basic OD principles suggests that this same process may also make a contribution to general aviation safety. Obviously, there are significant differences in the flight operations involved. While the same OD principles may be relevant in both areas, their specific application may vary widely.

One important difference is in cockpit organization. Airlines typically have a tightly defined, formalized structure in the cockpit. Each crewmember has certain responsibilities and tasks which are specifically delineated. This order works well in a controlled environment with a relatively small group and a single common and very specific goal. General aviation, on the other hand, is much less structured and much more diverse. Under these conditions, effective utilization of OD principles may be a greater challenge.

One potential application for these concepts is the definition of the optimum role for the graduate of a pinch-hitter course, such as that sponsored by the Aircraft Operators and Pilots Association (AOPA). Presently, the main objective of this course is to teach the passenger in the right seat how to land in an emergency situation should the pilot become incapacitated for any reason. This approach, however, is an underutilization of a resource with considerably greater potential. The pinch-hitter can improve the safety and the efficiency of the flight by helping with radio communications, opening charts, finding airports, scanning for traffic, etc. Almost invariably, pinch-hitters enjoy the course. Many of their fears about flying are allayed, and much of the mystery disappears. There is no reason a pinch-hitter cannot make a positive contribution to the flight even if his role consists of only elementary tasks such as switching frequencies on the radio.

The present pinch-hitter training course is similar to giving resource management training only to captains. Only half the job is done if a newly trained pinch-hitter gets back in the plane with available resources to offer the left seat pilot - only to discover that the left seat pilot is neither aware of the potential resource nor of ways to utilize it. The left seat pilot should recognize the potential of this new resource and needs guidelines to assist him in this way. Providing the needed help is a challenge for all of us who are concerned with GA training and safety.

Another potential GA application for OD has sometimes been called "the occasional copilot problem." It involves two qualified pilots, frequently, two flight instructors both at the control of the aircraft. This situation, seriously regarded by some people as quite dangerous, can create very real role perception problems. Having two pilots, neither clearly designated as pilot-in-command (PIC) is only slightly more hazardous than having both of them playing that role. (The airline version is flying with two captains.) In most situations, the PIC sits in the left seat. And often, this seating arrangement alone is enough to clarify the PIC and copilot. However, during instruction, the student usually sits in the left seat with the instructor as PIC in the right seat. This situation increases the vulnerability of instructor pilots to conflict when flying together, particularly under stressful conditions. Unfortunately, these are precisely the times when it is important to have no confusion regarding roles and exactly the situations under which the right seat instructor pilot is trained and conditioned to command the plane.

Another difference between the airline and non-airline environments is the technology available and the conditions under which it is used. Very few airline crews fly a B-727 one week and a B-737 the next. However, GA pilots frequently fly substantially different aircraft without anything resembling airline cockpit transition training. Even similar aircraft types may have radically different cockpit equipment.

One last difference between the airline and general aviation environments relates to the motivational bases of the employees. Because it is the nature of their job, airline flight crews are more highly motivated to work as a team for the safety and effectiveness of the operation. Thus, one would anticipate a minimal amount of conflict between crewmembers. It would be difficult to make such a generalization concerning general aviation. For example, the copilot may be brand new in a particular type aircraft. However, knowing that he has just completed two years in the left seat PIC (or even worse, as a flight instructor in the right seat) may shed some light on his motivation. He may be flying as copilot because he wants to build flight time in this particular type of aircraft in hopes of upgrading to the left seat. Or, he may be flying right seat because he feels more comfortable in that position. Either motivation could lead to conflicts with other crew members.

One problem for general aviation pilots is that no where do they receive training in resource management. The FAA provides outlines for flight training curricula and sets standards through flight and ground testing. None of the flight test guides for any of the five pilot certificates (student, private, commercial, flight instructor, and airline transport) mention such topics as resources, crew coordination, etc. Some flight training programs provide instruction on utilizing the right seat crewmember, particularly in light twin training. However, during the checkride, most candidates are expected to ignore any resources that might occupy the right seat.

Since the GA environment is neither static nor strictly ordered, and the pilots potentially are quite diverse in skill level, motivation, and experience, a model is needed that will allow the pilot to identify resources and utilize them to more efficiently and safely fly the aircraft. One such model, developed by Bobbitt and Behling (1980) for business management areas, provides a diagnostic approach which applies well to aviation resource management.

The diagnostic approach first classifies resources into four variables: task, technology, structure, and people. "Task" is defined by the objectives of the situation - these must be operationally meaningful statements. For example, in a situation where the landing gear will not extend, the task would be to keep the aircraft flying while examining options to remedy the problem.

"Technology" refers to the equipment in the aircraft. As mentioned earlier, it is not infrequent to have the same type aircraft equipped quite differently. This is a case of having different resources available. For the gear problem mentioned above, the presence of an autopilot or capable co-pilot could certainly change how the pilot would deal with a problem.

"Structure" refers to the communication and authority structure. Is the left seat pilot accustomed to, and most comfortable when flying alone? Does the right seat know to not initiate actions without approval or at least notification of the left seat? The two flight instructors with similar role expectations flying together is an example of the potential problem of the authority structure.

The last variable, "people", focuses on the attitudes, expectations, skills, and abilities of the pilots. A co-pilot with two hours time in a particular type aircraft is a different resource than one with 1000 hours time in type. Becoming acquainted with the capabilities of the person in the other seat should be a part of preflight planning. It should be noted, however, that the importance of differences may increase or decrease depending on the task. If the task is reading approach plates, for example, the numbers of hours in type is irrelevant.

It is important to realize that the four variables are mutually dependent on each other. A change in one will have an effect on another. Furthermore, once the diagnostic breakdown is completed, a pilot has been provided with little more than that - an identification of potential resources. Stopping at this point puts a pilot at the same level as the left seat pilot not knowing what to do with a pinch-hitter.

The second stage, therefore, of the diagnostic approach is to determine the utility of the four variables. Drawing from J. D. Thompson's (1976) "Three C's of Management", the variables are evaluated as constraints, contingencies or controls. Constraints are those variables over which a pilot has no control. Many of the technology variables are constraints. For example, if a transponder does not have an encoding function, then that is its limitation. The importance of constraints is recognizing that they cannot be changed to solve an inflight problem.

Contingencies are the "what-ifs", events which may or may not occur. A good resource manager excels in the handling of this area. It might mean, for example, assigning to the right seat the task of setting up navaids and frequencies for the approach. A contingency for which the PIC should be prepared is discovering that the wrong approach plate was used and then how to respond in redirecting resources.

Controls are those items over which the pilot does have command. The utility of the right seat will depend on how the people in both seats have learned to coordinate their activities through an authority structure. If the left seat is unsure of who is controlling what, performance of tasks will suffer.

The importance of this model lies in its adaptability to different situations. It is a model which fits the dynamic environment of general aviation. However, for this diagnostic approach to work, both left and right seat occupants must be aware of its limitations - namely, as a model which classifies. If the classification is faulty, then so may be the performance of tasks. This statement represents a fairly often heard criticism by left seat pilots of their experiences trying to utilize the right seat. A copilot frequently causes a workload increase (for the left seat) because not only must the aircraft be flown but now the right seat must be monitored as well. This situation is correctable using the diagnostic model. Monitoring is necessary but so is planning for the contingencies and coordinating activities among the crewmembers.

Utilization of OD principles can make a significant contribution to the general aviation sector. It provides an organized framework of problem-solving skills which can deal with the variances of skills, abilities, role perceptions, and motivations of the pilot populations involved. The skills developed must be appreciated by both right and left seat occupants and their utilization clearly understood prior to the flight. These concepts once developed and implemented should also be supported by the FAA and by all other leadership elements in GA.

REFERENCES

Beach, B. E. 1979. Line-oriented flight training. Proceedings of a NASA/Industry Workshop, Resource Management on the Flight Deck. NASA Conference Publication 2120.

Behling, O. and H. R. Bobbitt, Jr. 1980. Diagnosing and treating management problems. (Unpublished manuscript from the Ohio State University).

Thompson, J. D. 1976. Organizations in Action. New York, NY: McGraw-Hill Book Company.

VALIDATION OF A PROPOSED PILOT TRAINEE SELECTION SYSTEM

Jefferson M. Koonce
U.S. Air Force Academy, Colorado

The use of an Aircrew Psychomotor Test Device and the Air Force Officer Qualifying Test have been proposed for the selection of U.S. Air Force pilot training and candidates. Random sample of the class of 1978 and the class of 1979 were given the proposed tests and followed through undergraduate pilot training. The results cast serious doubt as to the utility of these tests in selecting U.S. Air Force Academy cadets for pilot training.

INTRODUCTION

With the current high costs of undergraduate pilot training (UPT) there is a need to accurately screen candidates so that those admitted into pilot training will have the highest probability of successfully completing the program. The ability to reduce the numbers who fail to complete pilot training to less than ten percent through an improved pilot trainee selection system is the goal of ground based screening (GBS). The proposed GBS system (McGrevy and Valentine, 1974) would use the Air Force Officer Qualifying Test (AFOQT) form N and an aircrew psychomotor test (APT) device to select pilot training candidates on cognitive factors (AFOQT) and perceptual-motor factors (APT). Gould (1978) reviewed the development of the AFOQT, which has been used by the U.S. Air Force since 1951 and has undergone fourteen revisions over the years. The AFOQT results are organized into five composite scores (officer quality, verbal, quantitative, pilot, and navigator-technical) based upon various combinations of the eighteen subtests. The full AFOQT takes approximately seven hours to administer and it may be given individually or as a group test. Gould also reviews the standardization of the present Form N of the AFOQT.

Passey and McClaurin (1966) provide a comprehensive review article on psychomotor testing over the years and suggest certain refinements for the updating of psychomotor test devices. Recent technological advances in solid state electronics has caused a renewed interest in the use of psychomotor tests in pilot selection. McGrevy and Valentine (1974) report on the development and validation of two psychomotor tasks presented by a psychomotor device (APT) which is reliable, easily maintained, and provides accurate measures of subject performance. The first of the tasks is a two-hand coordination pursuit task and the second is a complex coordination test requiring x-y compensatory tracking with a joystick and a single z-axis tracking using footpedals. McGrevy and Valentine (1974, pg. 9) state that "the addition of the psychomotor tests to the AFOQT can enhance the prediction of pilot training success."

The proposed implementation of a GBS system requires that it be used to select pilot training candidates from all sources; AFROTC, OTS and USAFA. Because the U.S. Air Force Academy's population is a highly stratified sample of the population at large and because it has a highly refined selection program along with four years of motivation and training, the utility of applying GBS to Academy cadets has been questioned. In fact, over the past ten years, the attrition of Academy graduates for flying deficiency in pilot training has generally been ten percent or less, the goal of the GBS system. Thus, before such a program as GBS is applied to a very select population the validity of that program for screening pilot candidates should be demonstrated as being as good or better than the existing system. The previously claimed validity of the AFOQT and the APT have been correlation coefficients between candidates' performances on the tests and their pass/fail performances in pilot training, not a demonstration of the relationship of their predicted performances with actual pilot training performance.

The purpose of this study was to demonstrate the relationship between the pilot training performances predicted by the GBS system and the candidates actual performance in pilot training. The experimental hypothesis was that GBS will significantly enhance the capability to identify those pilot training candidates who will fail to complete pilot training while not causing a significant number of false positives, that is, denying access to pilot training those who would have been successful. In demonstrating the validity of applying the GBS system to the Academy, it was decided that (1) pilot qualified cadets should take the AFOQT and APT, and be ranked in order of performance scores, (2) those failing to score above the proposed cut-off level be identified, (3) all candidates be permitted to go to pilot training, and (4) after completion of pilot training their pilot training performance would be compared with their scores on the AFOQT and APT.

METHOD

Apparatus

The Air Force Officer Qualifying Test (AFOQT), Form N, were furnished by the Air Force Human Resources Laboratory (AFHRL) for administration of the test to Academy Cadets. See Gould (1978) for detailed description of the AFOQT.

Two aircrew psychomotor devices manufactured by Systems Research Laboratories for AFHRL were positioned in individual rooms of the Behavioral Sciences Laboratories U.S. Air Force Academy. Each device automatically presented and scored two tasks. In the first task the subject was required to track a target with his cross using two hand controllers. The right-hand controller moved the cross horizontally and the left-hand controller moved the cross vertically. The target moved clockwise in a circular manner with minor variations in velocity. Both the target and the subject's cross were presented on a CRT and the controllers were to the left and right of the screen slightly above 'armrest' level. The second task presented cross hairs dividing the CRT screen in half vertically and in half horizontally. With a stick similar to that used to control aircraft the subject attempted to maintain his cross in the center of the cross hairs. Moving the stick to the right caused the cross to move left on the screen and moving the stick forward (away from the subject) caused the cross to move toward the bottom of the screen. While trying to keep the cross centered, the subject also had to keep a vertical bar, located on the bottom of the CRT, centered laterally upon the vertical cross hair with the use of footpedals. Pressing the right footpedal caused the vertical bar to move from right to left on the screen, and the left footpedal would move it to the right. The APT device was controlled by a cassette tape which gave instructions to the subject, controlled the presentation of the tasks, and caused the subject's error data to be collected and presented for the experimenter to record after each session. The total time for the APT device to present both tasks was eighteen minutes. See Sanders, Valentine, and McGrevy (1971) and McGrevy and Valentine (1974) for further details on the APT.

Subjects and Procedure

All subjects were U.S. Air Force Academy cadets randomly selected by the Academy's computer and given an appointment to be present for the testing. When the subjects arrived for their scheduled appointment they were briefed on the nature of the study, informed of their status as a volunteer, and asked to sign a consent form before participating in the study.

In December 1977 the AFOQT was administered to approximately two hundred randomly selected cadets from each class, 1979 through 1981. The testing was conducted on a Saturday in several large lecture halls with adequate lighting and writing surfaces. Cadets taking the AFOQT were excused from a Saturday morning training session and inspection. Administration of the AFOQT took approximately seven hours, not including the break for lunch.

In the Spring of 1978 those first classmen, cadets of the class of 1978, who were pilot qualified and had taken the AFOQT were asked to take the Aircrew Psychomotor Test in the Behavioral Science Laboratory. These subjects (N = 129) had either completed the T-41, light aircraft, flight training or had completed at least half of the program at the time of the APT session. Additionally, 261 pilot qualified second class cadets of the class of 1979 were randomly selected to take the APT. None of these cadets had taken T-41 flight training, however, a portion (8.8%) of them had had some flight training outside of the Academy environment. The second classmen who took the APT were a different sample from those second classmen who had taken the AFOQT.

When a cadet reported for the APT session he was told the nature of the study and the task to be performed and asked to sign a consent form volunteering to participate. Those not desiring to participate were dismissed.

RESULTS

AFOQT

The AFOQT criteria for acceptance to pilot training is at least the 25th centile on the pilot composite, the 10th centile on the navigator, and at least at a total of 50 when combining the pilot and navigator centile scores together. The data on the combined pilot and navigator scores are in Table 1. There was no significant differences in the performance of these two groups on the AFOQT. Using the AFHRL criteria for the AFOQT, 8.4% of those pilot qualified cadets in the class of 1978 who took the AFOQT would have been denied access to pilot training if GBS were implemented. Similarly, 13.6% of the class of 1979 would have been denied access to pilot training. The proportion of those who would have been denied access to pilot training who actually succeeded in pilot training was 80% for the class of 1978 and 81.25% for the class of 1979. The correlation between the AFOQT scores and UPT pass/fail were 0.149 and 0.260 for the classes of 1978 and 1979 respectively.

TABLE 1

AFOQT Scores and Pilot Training Performance

		AFOQT Scores			
		Class of 1978		Class of 1979	
		Failed	Passed	Failed	Passed
UPT	Passed	8	99	13	94
	Failed	2	10	3	8

APT

The Air Force Human Resources Laboratory has proposed that only the second psychomotor task on the APT be used for pilot screening because the first task does not contribute signficantly to the validity of GBS. Normative data on the first task was not available. The scoring of the second task is developed by first converting the raw error scores on x (stick horizontal), y (stick vertical), and z (rudders) axes to z-scores using the population means and standard deviations obtained earlier from the AFROTC and OTS population, and then summing the three z-scores. A negative sum would be better than average which should be zero. If a candidate should have a z-score sum of 3 or more that would eliminate him from consideration for pilot training.

There was no significant difference between the first and second classman on the APT. Had the APT been used to screen these candidates for pilot training using all of the data, 9.3% of the first classmen and 6.0% of the second classmen would have been denied access to pilot training (Table 2). The proportion of those who would have been denied access to UPT based on their APT scores but who actually passed pilot training was 83.33% for the class of 1978 and 91.67% for the class of 1979. The correlations between the APT scores and UPT pass/fail were -0.036 and -0.086 for the classes of 1978 and 1979 respectively.

TABLE 2

Aircrew Psychomotor Test Scores and Pilot Training Performance

		Aircrew Psychomotor Test			
		Class of 1978		Class of 1979	
		Failed	Passed	Failed	Passed
UPT	Passed	10	107	11	168
	Failed	2	10	1	20

DISCUSSION

Neither the AFOQT or the APT scores discriminated very well between those cadets who passed pilot training and those who failed to complete pilot training. In fact the distributions of the scores of those who passed and failed pilot training were very similar. The lack of significant correlations between the test scores and pilot training performance supports this observation.

The cadets in the class of 1978 had completed or were halfway to completion of a light aircraft training program while those in the class of 1979 had not yet entered the light aircraft training program. On the AFOQT the class of 1978 scored slightly higher than the class of 1979 but insufficient for statistical significance. One would expect such an increase in score due to an additional year of college education, but the addition of light aircraft training did not add sufficiently for a difference between classes. Surprisingly, the second class cadets, class of 1979, performed slightly better than the class of 1978 on the Aircrew Psychomotor Test device; however that difference was not enough to reach statistical significance. The second APT task was very similar to flying an ILS approach with a flight director, but on one of the axes the control inputs were reverse those of an aircraft while on the other axis it was the same. This might have been a source of confusion and resulting higher error for the subjects in the class of 1978 since they had more flight experience than those in the class of 1979.

In general, the use of both the AFOQT and the APT would not have reduced the percent of attrition of Academy cadets in UPT. But, the prospects of denying access to pilot training to a group of cadets of which at least eighty percent would probably be successful in pilot training may have some serious side effects on morale and the willingness of the individuals to complete their education at the Academy. Many cadets come to the Academy with a very strong desire to become an Air Force pilot and by the time they are first or second classmen they represent a sizeable investment.

One alternative suggested has been to administer these tests to the cadets when they first come to the Academy, summer prior to the freshman year, or even before they arrive at the Academy. What that might do to the potential pool of future Academy candidates is unknown, but the effects of temporal decay of five years on the predictive capability of these tests would probably render them even less useful than they were for the classes of 1978 and 1979, college seniors and juniors.

At present, the Air Force Officers Qualifying Test and the Aircrew Psychomotor Test do not seem to be of any significant value for the screening of Air Force Academy cadets for undergraduate pilot training.

REFERENCES

Gould, R. B. Air Force officer qualifying test form N: Development and standardization. Brooks AFB, Texas: Air Force Human Resources Laboratory, Technical Report AFHRL-TR-78-43, August 1978. (AD 732 210)

McGrevy, D. F. & Valentine, L. D., Jr. Validation of two aircrew psychomotor tests. Brooks AFB, Texas: Air Force Human Resources Laboratory, Technical Report AFHRL-TR-74-4, January 1974.

Passey, G. E. & McClaurin, W. A. Perceptual-psychomotor tests in aircrew selection: Historical review and advanced concepts. Lackland AFB, Texas: Air Force Personnel Research Laboratory, Technical Report PRL-TR-66-4, June 1966. (AD 636 606)

Sanders, J. H., Valentine, L. D., Jr., & McGrevy, D. F. The development of equipment for psychomotor assessment. Lackland AFB, Texas: Air Force Human Resources Laboratory, Technical Report AFHRL-TR-71-40, July 1971. (AD 732 210)

The opinions expressed in this article are those of the author and do not necessarily reflect official opinion of the U.S. Air Force or the U.S. Air Force Academy. The author would like to express his sincere appreciation to the members of the Department of Behavioral Sciences, U.S. Air Force Academy, who assisted in the administration of the AFOQT and the APT.

SEX AS A MODERATOR VARIABLE IN THE SELECTION AND TRAINING OF PERSONS FOR A SKILLED TASK

THOMAS M. McCLOY and JEFFERSON M. KOONCE

United States Air Force Academy
Colorado

The emergence of females into traditional male roles in industries as well as the military has caused concerns over whether it might require modification of the selection and training procedures that have been developed previously to optimize the screening of males for those roles. Natural curiosity has asked how the sexes differ in physical strength capabilities, but the effects of any differences in cognitive styles or psychomotor abilities have not been fully explored.

For many years paper and pencil tests have been used by the military services in the screening of candidates for most officer training and flight training programs. Gould (1978) reviews the development of the Air Force Officers Qualifying Test (AFOQT) which has been used by the U. S. Air Force since 1951 and has undergone fourteen revisions in the past twenty-seven years. The test results are organized into five composite scores (officer quality, verbal, quantitative, pilot, and navigator-technical) based upon various combinations of the eighteen subtests which include electrical maze tracing, rotated blocks, and aerial landmarks.

Psychomotor ability has been used off and on as a predictor of pilot training success since 1919. The use of psychomotor tests in the U. S. Air Force was discontinued in the early 1950's because of difficulties in calibrating and maintaining the machines under decentralized testing conditions. McGrevy and Valentine (1974) report on the development and validation of two psychomotor tasks presented by a solid state psychomotor test device (PTD) which is reliable, easily maintained, and provides accurate measures of subject performance. The device automatically presents instructions to the subjects, gives them practice, and then tests them on two tasks: a two-hand coordination pursuit task (PTD1) and a complex stick-and-rudder coordination task (PTD2). McGrevy and Valentine state that "the addition of the psychomotor tests to the AFOQT can enhance the prediction of pilot training success." (pg.9).

This paper tells of a series of studies designed to determine just how different are males and females in cognitive styles and psychomotor abilities, whether one could expect a gender difference in the acquisition of skills, and whether or not prediction equations developed to optimize the selection of males for a skill would be adequate for the selection of females for those same skills. Also, once the males and females were trained to a given skill level on a basic task, would they exhibit equal transfer of training to more complex tasks? The research presented here used a battery of cognitive and psychomotor tests similar to those proposed by McGrevy and Valentine for the Air Force, proceeded to measure the performance in the acquisition of basic flight skills, and noted the effects upon transfer of training to complex flight skills. The cognitive tests were the Identical Pictures Test (perceptual speed), Map Memory Test (visual memory), Cube Comparison Test (spatial orientation), and Maze Tracing (spatial scanning) from the Ekstrom, French, Harman, and Derman (1976) battery, and the Embedded Figures Test (field independence) from the Educational Testing Service (Witkin, Oltman, Raskin, and Kark, 1971). Three psychomotor tests were administered: the pursuit rotor task (PR), the PTD1, and the PTD2.

STUDY 1

The subjects were fifty-one male and fifty-two female volunteer fourth class (freshman) cadets at the U. S. Air Force Academy. Initially they participated in three 50-minute sessions over a two month period.

The first session consisted of small group administration of the first four cognitive tests which were quite similar to subtests forming the pilot scale of the AFOQT. In the second session the subjects were individually given the EFT and the three psychomotor tests. In the third session each subject was instructed in how to perform four basic flight maneuvers (climb, cruise, descent, and turn) using a desk-top flight simulator, given practice on the maneuvers, and then tested on his or her performance while flying these maneuvers in simulated smooth and turbulent air conditions.

The performance of the males and females on the cognitive tests were in the directions anticipated from general population data, in that the females were somewhat faster on the IP test, while the males tended to be better on the MM, CC, and MT tests. Also, the males were faster on the EFT test (more field independent), but not significantly faster. Only on the CC test was the sex difference statistically significant ($p < 0.001$). On all of the psychomotor tests and all of the simulator flight maneuvers in rough and smooth air the males were significantly better than the females ($p < 0.001$). The lack of significant differences between the sexes on the cognitive tests might be due in part to the selection process for cadets at the Academy which tends to result in cognitively more homogeneous groups than would be expected in the general population.

The scores on the cognitive and psychomotor tests were used as predictor variables in a stepwise multiple regression technique to predict the flight maneuver performance in smooth air, rough air and overall for males and females separately. For the males, the PTD2 and the IP tests were the significant predictors; but for the females, the PR and CC tests were the significant predictors. The multiple Rs ranged from0.388 to 0.526, all statistically significant ($p < 0.05$). The multiple Rs in predicting flight performance for the males and females together did not differ significantly from those for either sex separately. The great dissimilarities in the equations for males and for females tend to support the notion that to select those females with the greatest probability of success in pilot training one should develop the weights for the predictor tests from females in the population as opposed to using those weights derived from males in the population.

STUDY 2

A year later those subjects who were still at the Academy (45 males and 45 females) were asked to return for two or three additional fifty-minute sessions. Because such a significant sex difference in the performance of the four maneuvers had been observed we asked whether it was attributable to sex differences in cognitive interpretations of the tasks or motor translations of the desired responses. The subjects were given a single trial pre-test on the four maneuvers. Based upon the pre-test performances they were divided into three equal groups of males and females for control (CON), cognitive training (COG), and cognitive and motor training (COG+MOT). The COG training consisted of verbal interpretation of eight different flight attitudes with verbal responses as to how one should respond with the throttle and yoke to bring the aircraft to a desired position. The COG+MOT training required the same verbal interpretation of the same eight attitudes, but instead of telling how one should respond to bring the aircraft to a desired position they actually made the initial responses to the throttle and yoke. To prevent practice, the COG+MOT subjects were not allowed to fly all the way to desired position, only make the initial input to the controls. Both experimental groups were given feedback at the end of each of the eight trials.

All three groups, CON, COG, and COG+MOT, were then tested on their abilities to perform the four basic flight maneuvers, and then they continued to practice them until their performance reached a preset criteria. The dependent variables were the performance on the first trial and the total trials required to meet criteria. The first trial performance indicated a difference between the sexes ($p < 0.02$) but no differences between groups. On the other hand, the trials to criteria reflected a significant difference between the performances of males and females ($p < 0.0001$) and a significant difference between groups ($p < 0.10$); see Figure 1. The multiple regression equations became stronger as the subjects overall were broken down into their respective training conditions, and even more when divided into sex within groups:

Overall	Group	Sex Within Group
0.593	CON 0.652	M 0.850
		F 0.782
	COG 0.783	M 0.733
		F 0.871
	COG+MOT 0.615	M 0.394
		F 0.763

These results indicated that sex is indeed a moderator variable in the prediction of basic flight performance. Also, novices who are not very familiar with a skilled task to be performed will benefit far more from both cognitive and motor training than from cognitive training alone. Perhaps once the subjects become more skilled the cognitive training will be just as effective as cognitive and motor training.

STUDY 3

The remaining subjects, 43 male and 38 female cadets, were then instructed on how to perform a chandelle, and then they practiced it until their performance met predetermined criteria. A chandelle is a maximum performance climbing turn that is more cognitively demanding than the four basic maneuvers previously learned. Prior to the learning of the chandelle all of the subjects were equally skilled in the performance of the four basic maneuvers. No sex differences in the learning of the chandelle were anticipated. The trials to criteria on the chandelle reflected a significant difference between the sexes ($p = 0.056$) and a significant difference based upon the previous group training ($p < 0.02$); see Figure 2. Separating the cognitive tests from the psychomotor tests in the prediction of all of the subjects performance on the chandelle, the multiple R for cognitive predictors was 0.414 while for psychomotor predictors it was 0.225. This indicates the more cognitive nature of the chandelle maneuver. In the prediction of trials to criteria on the chandelle overall, by group, and by sex within group the results were similar to those of the four maneuvers:

Overall	Group	Sex Within Group
0.488	CON 0.589	M 0.980
		F 0.690
	COG 0.704	M 0.825
		F 0.971
	COG+MOT 0.684	M 0.889
		F 0.805

The results of this study continue to support the idea of sex as a moderator variable in the prediction of complex flight skills. The individual regression equations by sex had significantly different predictor variables which suggests that regression equations developed on one sex may not be optimum for selection of persons of the other sex. The fact that a sex difference persisted on the chandelle after all subjects were performing at the same level on the four maneuvers leads one to think that the males and females might differ in their abilities to transfer their training on these types of tasks. But, because the chandelle was so cognitively demanding relative to the four basic maneuvers we might not be seeing the phenomena of transfer of training as much as the learning of an entirely different type of task.

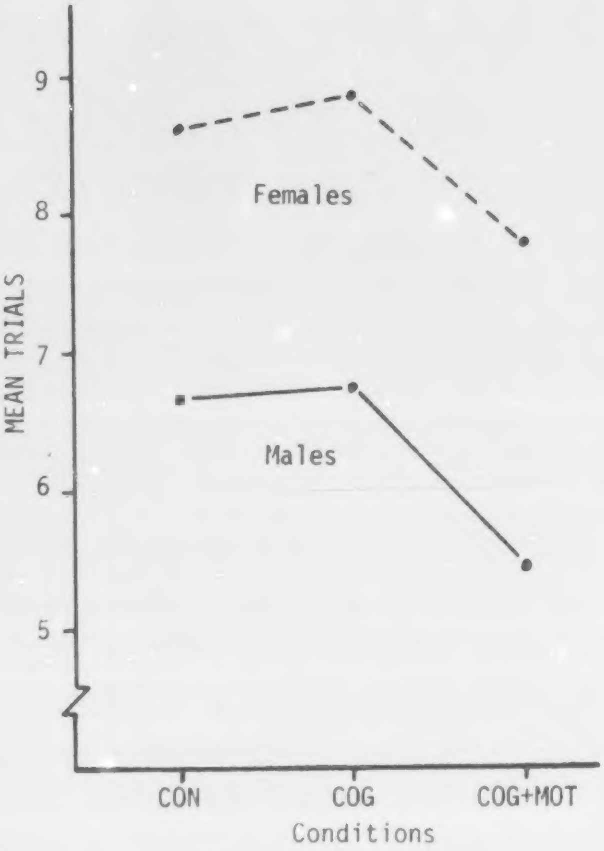


Figure 1. Mean trials to learn four maneuvers.

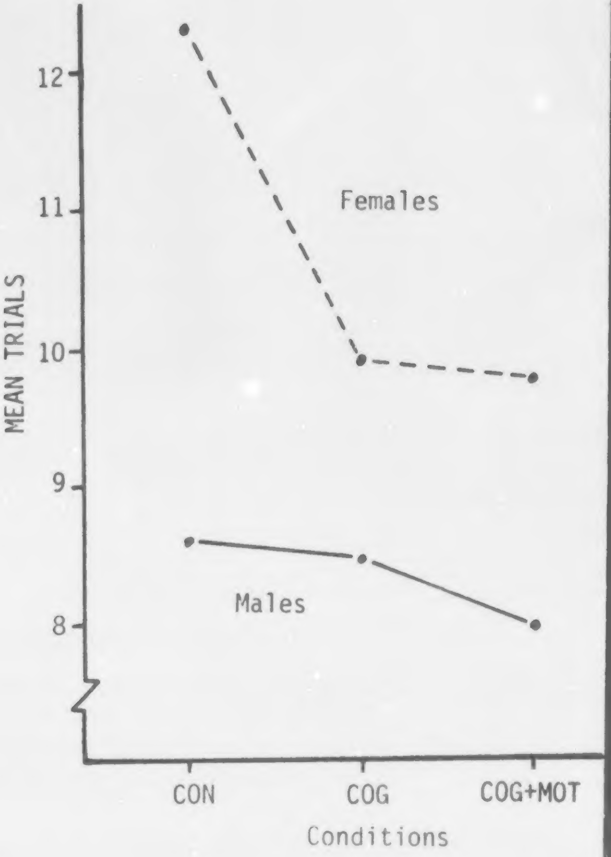


Figure 2. Mean trials to learn a chandelle.

Another study is being conducted this year at the Air Force Academy to see if the differences between the males and females can be completely "trained-out," and using comparable tasks will males and females transfer their training equally from one task to another.

REFERENCES

Ekstrom, R. B., French, J. W., Harman, H. H., and Derman, D. *Manual for kit of factor-referenced cognitive tests*. Princeton, New Jersey: Educational Testing Service, 1976.

Gould, R. B. Air Force officer qualifying test form N: Development and standardization. Brooks Air Force Base, Texas: Air Force Human Resources Laboratory, Technical Report AFHRL-TR-78-043, August 1978.

McGrevy, D. F. and Valentine, L. D. Validation of two aircrew psychomotor tests. Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, Technical Report AFHRL-TR-74-4, January 1974.

Witkin, H. A., Oltman, P. K., Raskin, E., and Karp, S. A. *Manual for the embedded figures test*. Palo Alto, California: Consulting Psychologists Press, 1971.

CHANGES IN THE US ARMY AVIATOR SELECTION AND TRAINING PROGRAM

by
W. R. Brown, J. A. Dohme and M. G. Sanders
US Army Research Institute
Fort Rucker, Alabama

Shortly after World War II the Army began to build its own aviation program. Early on it was recognized that the Army's training needs and those of the other services were sometimes disparate. One major disparity is the Army's Warrant Officer aviator program. The Army trains both Commissioned Officer and Warrant Officer aviators. While the officer flight training program was and is very similar to the flight training programs of the Air Force and the Navy, the Warrant Officer (WO) program is unique. Basically, the WO program permits individuals who have only a high school education to be trained as aviators.

The selection process is necessarily different for the two training programs. This paper will present an overview of those training programs with a focus on aviator selection; past, present and future.

The fact that the Army has two basic entry programs creates some special selection and training problems. The backgrounds of the individuals who enter flight training via these two avenues appear to be somewhat different. The majority of WOCs entering the program have prior enlisted service but a sizable come "off the street", i.e., civilian entry. The Commissioned Officers (COs) entering the program generally differ in two ways from the WOCs: they have more formal education and they have been through at least one "screening program" designed to eliminate those who do not have the potential to become officers. That is, all COs have undergone basic officer training in one of these three programs: ROTC (Reserve Officer Training Corps), OCS (Officer Candidate School), or the US Military Academy.

Training Rates

Over the years, the number of WOCs and COs trained has changed drastically. In the early days, (forties and early fifties), when the Army had more fixed wing (F/W) than rotary wing (R/W) aircraft, many more COs than WOCs were trained. The general approach at that time put COs into F/W aircraft and WOs into R/W aircraft. As the tactics and doctrine of the Army changed, going more to R/W than F/W, the number of CO and WO aviators trained has become more nearly equal. With these changes, the general rule of training COs primarily for F/W and WOs for R/W has also changed. All flight students now receive Initial Entry Rotary Wing (IERW) training as the first step in their flying career. Later, some are given advanced aircraft transitions, either F/W (T-42, U-21, OV-1, etc.) or R/W (CH-47, AH-1, UH-60, etc.). The number of CO and WO IERW graduates by year are shown in Table 1. This table reflects that there have been substantial changes in output from the IERW flight training program.

INSERT TABLE 1 ABOUT HERE

The increased training requirements observed in recent years came about for three primary reasons: (1) the loss of a large number of aviators trained during the Vietnam era, (2) the low retention rate of aviation WOs; only 46% of the aviation WOs trained during fiscal years 1976, 7T, and 1977 remained in the Army at the end of their three year obligated tour, and (3) increased aviation force requirements for the 1980's.

Attrition Rates

The attrition rates for both CO and WOC IERW students can be seen in Table 2. Prior to 1977, WOC flight training students were given OCS type training concomitantly with flight training. As Table 2 illustrates, the attrition rate for WOC students was quite high for the FY 75 and 76 timeframe.

INSERT TABLE 2 ABOUT HERE

An investigation into these attrition rates revealed that the majority of these attritees were being eliminated for non-flight related deficiencies, i.e., military or officer development deficiencies. This attrition was occurring during flight training, after considerable time and expense had been invested. In an effort to reduce costs in aircraft and time, a formal military development course was established to train WOCs in these military and officer skills prior to actual flight training. This course is called the Warrant Officer Military Development Course (WOCMDC) and is six weeks in length. Table 2 illustrates the percentage of attrition associated with IERW training both before and after the establishment of the WOCMDC. The bar graphs for the years 1977 through 1980 show that the WOC eliminations for flight related deficiencies closely match those of the officers. Most of the attrition takes place during the WOCMDC prior to flight training thereby reducing training costs. As noted earlier, the COs entering flight training have already been through this type of military development screening process, therefore attrition observed for COs in IERW is primarily due to flight deficiencies.

Flight Aptitude Selection Test (FAST)

The first operational selection instrument for the flight training program evolved from the development of two separate test batteries, one designed for the selection of officers as fixed wing trainees and one designed for the selection of Warrant Officer Candidates (WOCs) as rotary wing trainees. As a result of a field evaluation, the two were consolidated into one test battery in 1966. However, the resultant test, the Flight Aptitude Selection Test (FAST) still retained two forms; the Officer Battery (FAST-OB) and the Warrant Officer Battery (FAST-WOCB). Table 3 presents a listing of the components of the two batteries. These subtests can be grouped into four basic content areas -

(1) biographical data and interest information, (2) spatial ability, (3) mechanical ability, and (4) aviation information.

INSERT TABLE 3 ABOUT HERE

Each FAST battery resulted in three scores: (1) rotary wing aptitude score, (2) fixed wing aptitude score, and (3) a composite score. The composite score was used for selection purposes. The FAST maximum composite score is 519 for the WOCB and 562 for the OB.

FAST composite score medians (50th percentile) of 256 (SD = 60) and 300 (SD = 30) respectively for the OB and WOCB were found for 1977 through 1979 applicants. The minimum score required for entry into the flight training program was 155 for officers and 300 for WOCs. Tables 4 and 5 present the composite score means, by year, for officers and WOCs entering flight training. It can be seen from these tables that the composite score means for the students actually selected for flight training are considerably higher than the minimum. This is due to the large applicant pool from which the Army may select. The Army has the opportunity to select individuals with higher FAST scores and still fill the throughput/graduation requirements. This favorable selection ratio does, however, appear to be declining. Tables 4 and 5 indicate that with the increase in required throughput (reference Table 1) the composite score means for IERW students show a slight decrease over the last few years. It should be noted, however, that the average composite score for WOCs entering the program in 1981 is 324 which is at the 68th percentile score.

INSERT TABLES 4 & 5 ABOUT HERE

When using any selection instrument, one of the important factors to be determined is its predictive validity. Several predictive validity estimates have been made comparing composite scores on the FAST with overall flight grades. The most recent study (Eastman and McMullen, 1978a) found that the predictive validity coefficients of the FAST to overall grade were .38 for the WOCB and .44 for the OB. The estimates were adjusted for restriction of range on the predictor variable. Eastman and McMullen (1978a) also found that successful officer trainees obtained significantly higher FAST composite scores (278 vs 244) than eliminated officers ($t = 2.39$; $p<.05$). However, the FAST score difference between successful (mean = 342.6) and nonsuccessful WOCs (mean = 333.7) was not statistically significant.

Two factors probably account for the lack of statistical difference observed in the FAST scores of successful and nonsuccessful WOCs: (1) the highly restricted range of FAST scores for WOCs entering the flight training program, i.e., individuals selected for the WOC program scored in the 70-80th percentile range on the FAST, and (2) a large percentage of the WOCs were eliminated for deficiencies not related to flight aptitude.

Revised FAST

Several significant changes have occurred since the FAST was developed in 1966, e.g., the nature of the mission has changed to reflect the emphasis on rotary wing flight, especially Nap-of-the-Earth (NOE) flight. Additionally, the aircraft coming into the Army inventory have grown in complexity, and the applicant population has changed to reflect increased minority and female representation. Also, there were some difficulties with the FAST itself. It was composed of two rather lengthy batteries which were difficult and time consuming to score. Additionally, there was an unacceptable error rate at the Armed Forces Examination and Entrance Station (AFEES), in the administration and scoring of the two batteries (Eastman and McMullen, 1978b).

Based on these and other considerations, a Revised FAST (RFAST) was developed which became operational in March 1980. The RFAST has several advantages over the old FAST: (1) it contains only 200 items, chosen primarily from those items on the old FAST found to be most predictive of IERW performance, (2) it is easier to score, (3) the same test is given to both WOC and CO applicants.

Table 6 shows the seven subtests contained in the RFAST as compared to the original twelve subtests found in the FAST-OB and FAST-WOCB. A preoperational study was performed to obtain an estimate of the predictive validity of the RFAST relative to the IERW overall grade. A sample of 178 WOC students, previously selected based on scores from the old FAST, were given the RFAST shortly after entering flight training. This sample had a mean composite score of 112 with a standard deviation of 15. The predictive validity coefficient was found to be .33, corrected for restriction in range. Preliminary data on 2,517 applicants tested with the RFAST indicates a mean of 98 and standard deviation of 20.5. An initial cut score of 90 (35th percentile) was established for the RFAST. This cut score will allow Army selection officials to adjust to varying selection ratios and select the most qualified individuals available. A validity estimate for the officers has not yet been accomplished. It is anticipated that a full predictive validity effort will be completed in FY 82.

INSERT TABLE 6 ABOUT HERE

Current Research Efforts

Several other projects are in progress and show some promise for increasing the effectiveness and broadening the scope of the Army aviator selection program. A brief description of some of these projects is provided:

(1) FAST Fairness Evaluation - The old and RFAST batteries are being evaluated relative to the issue of fairness. Greater minority and female representation in the flight school applicant population dictates the evaluation of all forms of the FAST to insure equivalent predictive capabilities for all applicant groups. Research along these lines is being conducted in-house as well as through contractual support.

(2) Revised FAST Alternate Form - Army policy changed in 1980 such that one retest is allowed on the Revised FAST six months after original testing. The preliminary stages of research have been completed in the process of developing an equivalent alternate form of the RFAST. The development of the alternate form will allow the Army to properly implement the retest provision.

(3) Differential Mission Assignment of Aviators - At the current time, the Army graduates aviators in a dual track IERW training program in which approximately 25% of graduates earn their wings as Aeroscouts flying the OH-58 helicopter and the majority earn their wings as Utility aviators in the UH-1. ARI has been tasked to develop an assignment procedure to identify the aptitudes and abilities of student pilots in order to optimally assign them to one of four training tracks corresponding to the four basic helicopter missions: Aeroscout, Attack, Cargo and Utility. The mission analysis is nearing completion and work is beginning on the development of a test battery which will assess mission specific abilities and aptitudes.

(4) Performance-Based Aviator Selection System (PASS) - The PASS is a job sample test which evaluates the appliant's psychomotor and cognitive capabilities in the performance of actual flight maneuvers. The PASS program involves five one hour blocks of automated flight instruction and evaluation in a UH-1 helicopter flight simulator. The applicant progresses from the most basic aspects of helicopter flight up to complex high workload flight maneuvers (a VOR approach with a concurrent on-board emergency procedure). After validation, the PASS will be used as the second part of a two-step sequential selection program. The first step will be the RFAST which will be used for initial screening followed by a finer aptitude assessment with the PASS.

REFERENCES

1. Eastman, R. F. and McMullen, R. L. *The current predictive validity of the flight aptitude selection test.* US Army Research Institute Fort Rucker Field Unit, Research Memorandum 78-2, March 1978 (a).

2. Eastman, R. F. and McMullen, R. L. *Item analysis and revision of the flight aptitude selection tests.* US Army Research Institute Fort Rucker Field Unit, Research Memorandum 78-4, April 1978 (b).

3. Kaplan, H. *Prediction of success in Army aviation training.* US Army Research Institute Technical Research Report 1142, June 1965.

TABLE 1

COMPARISON OF THE NUMBER OF COMMISSIONED & WARRANT OFFICER GRADUATES FROM IERW FLIGHT TRAINING: FY 75 THROUGH FY 81.

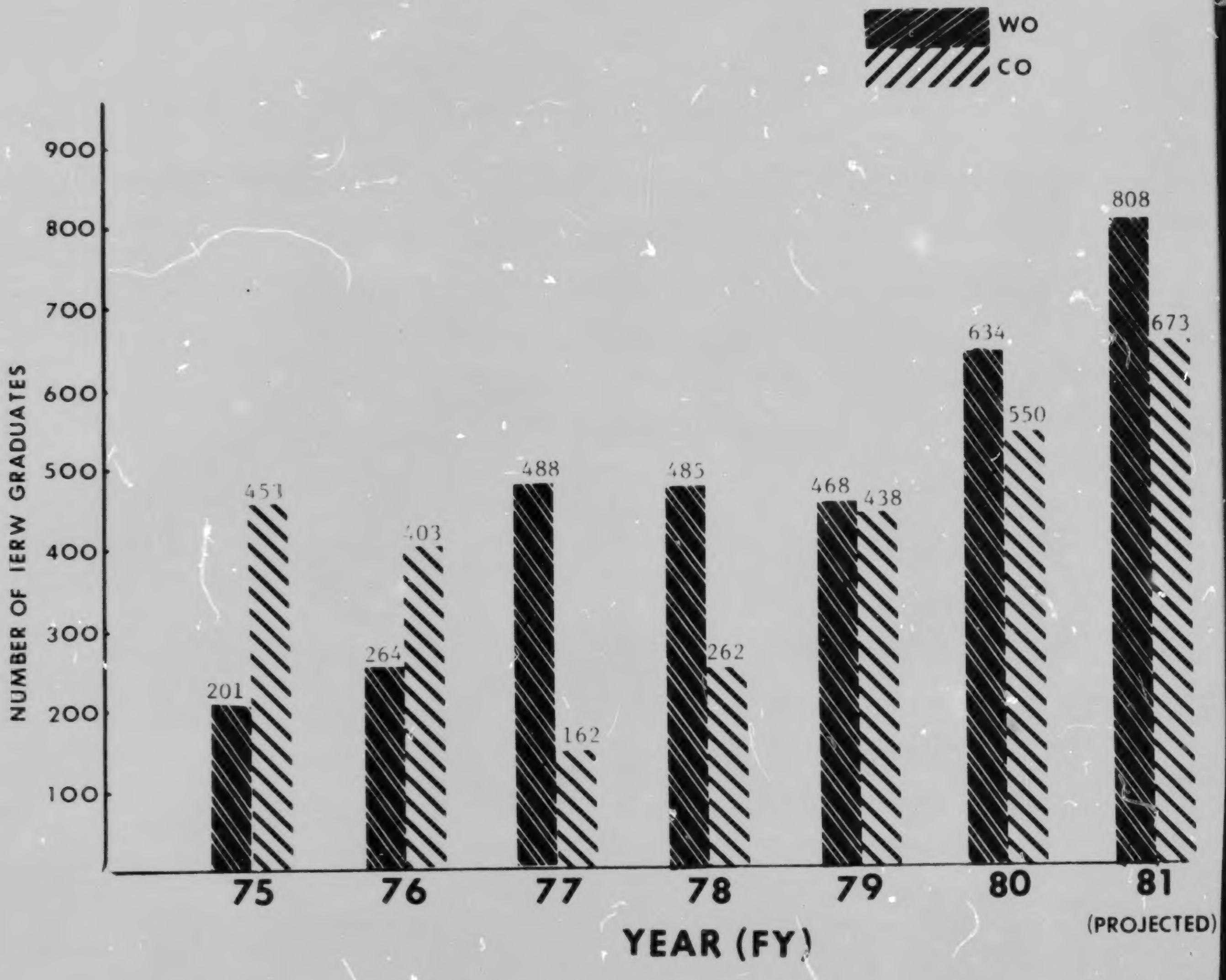


TABLE 2

IERW ATTRITION RATES FOR WOCMDC, WO FLIGHT TRAINING, AND CO FLIGHT TRAINING

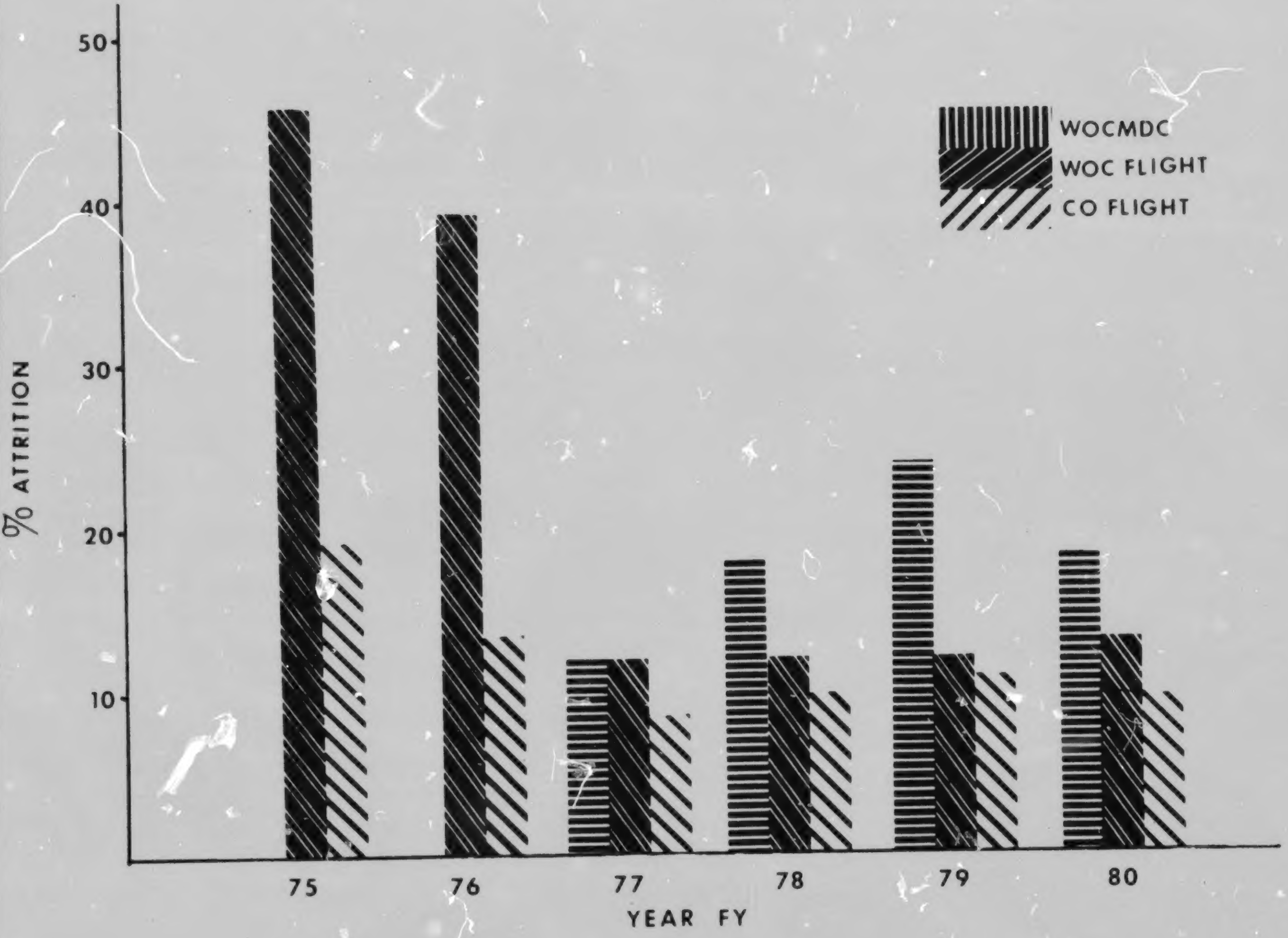


Table 3

COMPOSITION OF FAST BATTERIES

Test	Officer	Warrant Officer
Biographical Information	X	
Mechanical Principles	X	
Flight Orientation	X	
Aviation Information	X	X
Helicopter Information	X	X
Mechanical Information	X	X
Mechanical Functions	X	X
Visualization of Maneuvers	X	X
Instrument Comprehension	X	X
Complex Movements	X	X
Stick and Rudder Orientation	X	X
Self-Description		X

TABLE 4

MEAN FAST SCORES FOR WOCs ENTERING IERW

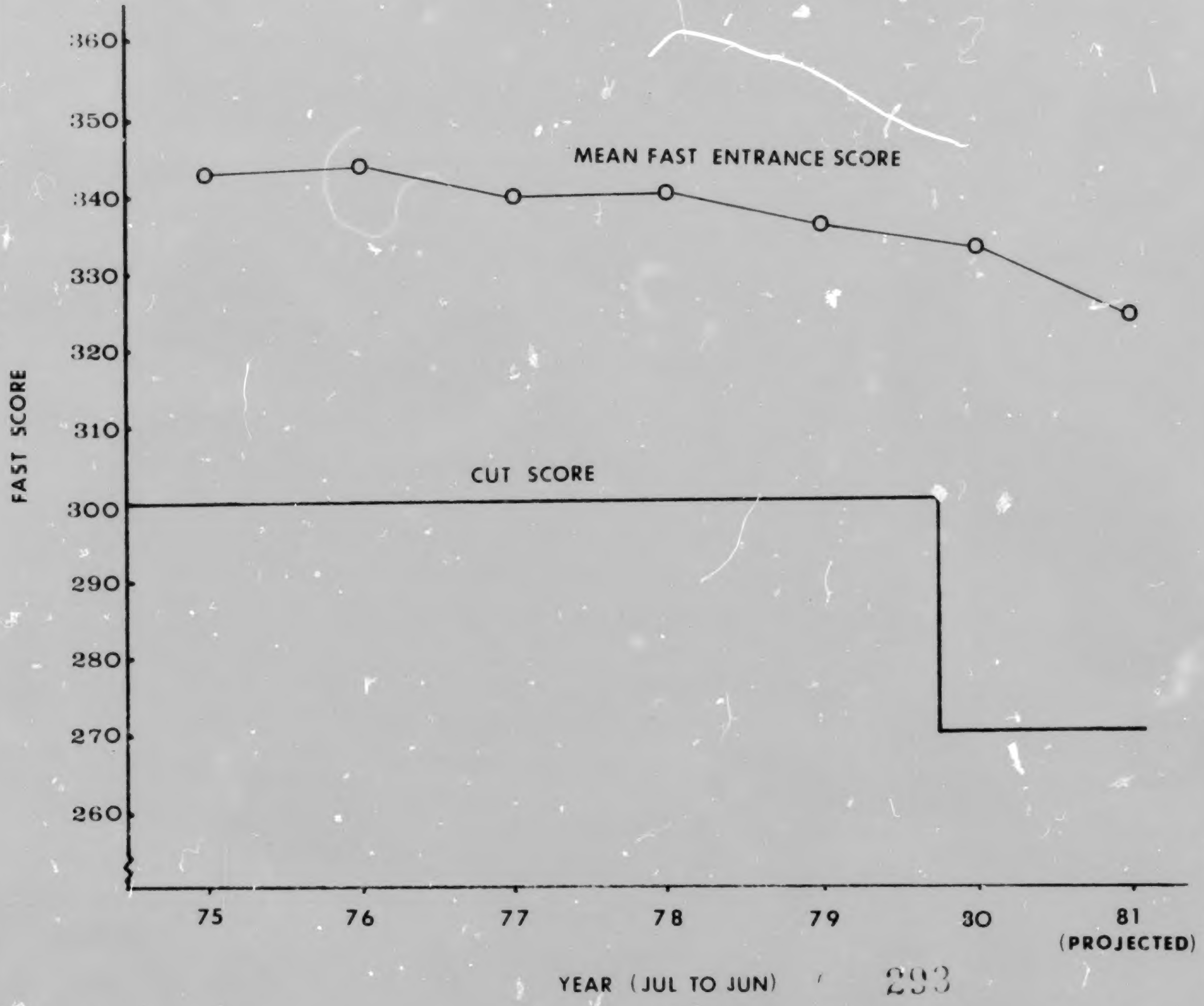


TABLE 5

MEAN FAST SCORES FOR COs ENTERING IERW

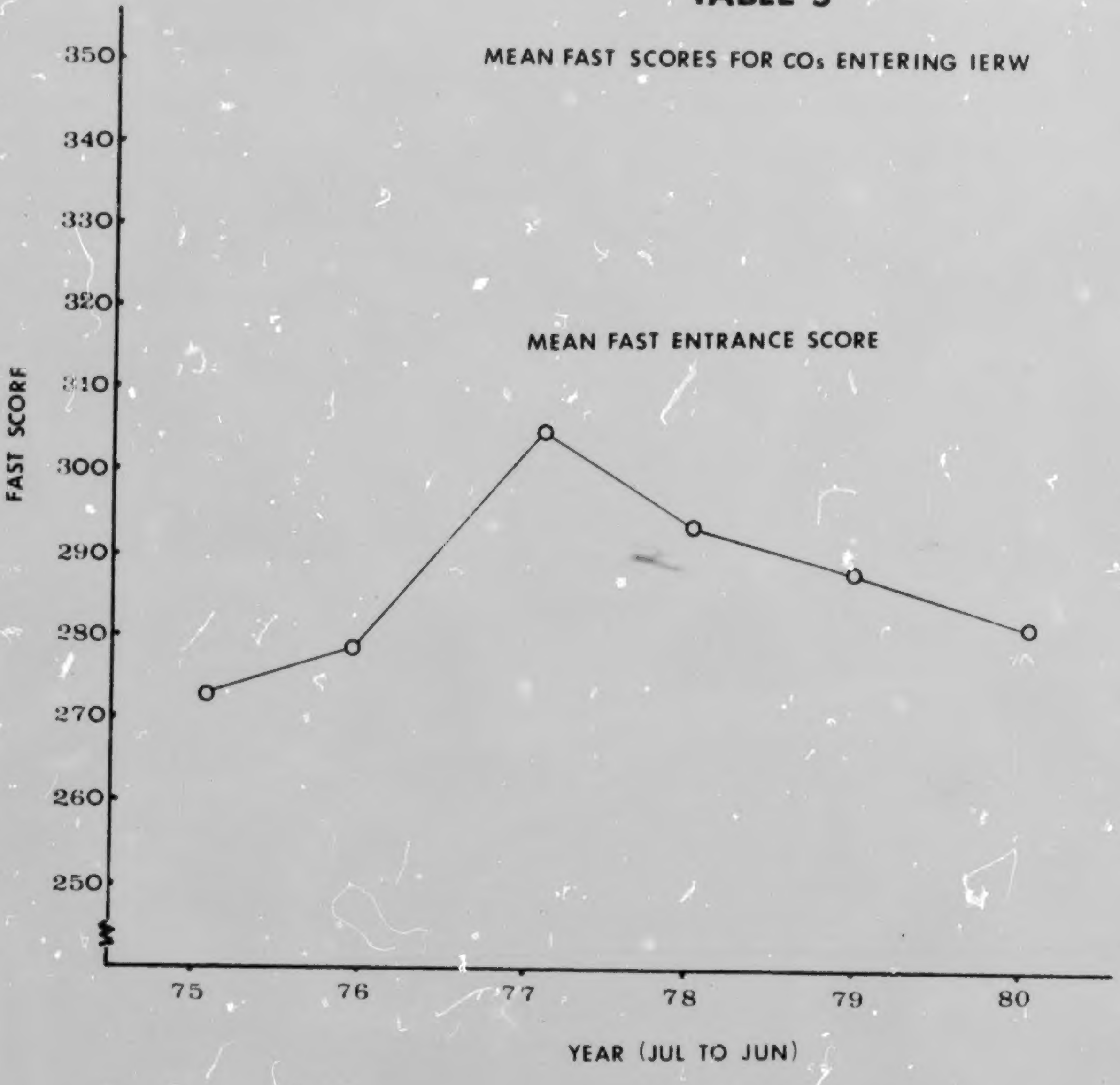


TABLE 6

COMPOSITION OF FAST BATTERIES

	Old FAST		Revised FAST
Test	Officer OB	Warrant Officer WOCB	Officer & WOC
Biographical Information	X		X
Mechanical Principles	X		X
Flight Orientation	X		
Aviation Information	X	X	
Helicopter Information	X	X	X
Mechanical Information	X	X	
Mechanical Functions	X	X	
Visualization of Maneuvers	X	X	
Instrument Comprehension	X	X	X
Complex Movements	X	X	X
Stick and Rudder Orientation	X	X	X
Self-Description		X	

INDIVIDUAL DIFFERENCES IN MULTI-TASK RESPONSE STRATEGIES

DR. DIANE DAMOS
THOMAS SMIST

Department of Industrial Engineering
SUNY at Buffalo
Buffalo, New York

ABSTRACT

This paper presents the results of two experiments examining individual differences in multiple-task performance. In the first experiment, the subjects were classified according to the response strategy they used to perform two discrete tasks. The subjects then performed three different task combinations on four successive days. A multivariate analysis of variance revealed significant between-strategy group differences in multiple-, but not single-, task performance. The second experiment was designed to determine if the results of the first experiment reflected between-group differences in information processing. The analyses indicated that one strategy group processed multiple-task information significantly more poorly than the other two groups.

INTRODUCTION

One of the major problems confronting aviation psychology today concerns excessive pilot workload. In some cases, recent advances in technology, which have greatly extended the capabilities of the aircraft, have put an increased demand on the pilot's limited information processing resources. In other cases, the nature of the mission itself, such as air-to-air combat, puts an extreme load on the pilot's processing capacities. Although some progress has been made in workload reduction through improved controls and displays, such changes do little to reduce the load inherent in some types of missions. Therefore, it seems important to select individuals who can deal effectively with high information processing loads.

To date, there has been little success in identifying individuals who can perform well under multiple-task conditions. However, Damos and Wickens (1980) found evidence of such differences when subjects were classified according to the response strategy they used to perform a discrete information processing combination consisting of a memory and a classification task. Three distinct strategies were identified: a simultaneous response strategy, an alternating strategy, and a massed strategy. A simultaneous response strategy is one in which the subject responds to both stimuli within some arbitrarily small interval (less than 100 msec). An alternating strategy is one in which the subject alternately makes one response to each task. If more than two responses are made consistently to one task before switching to the other, the strategy is classified as massed. Damos and Wickens found that the response strategy groups differed significantly on the memory-classification combination, but not on any single-task measure of performance. Additionally, the groups also differed in their performance on a tracking-tracking combination, but again no significant differences were found on any measure of single-task tracking.

This pattern of results may indicate that the response strategy reflects individual differences in multi-, but not single-, task performance. This paper presents the results of two experiments. The first examined whether there are consistent between strategy-group differences in multiple-task performance. The second determined whether the observed differences reflected fundamental differences in information processing or simply the effect of the strategy itself (a more detailed description of the experiments is given in Damos and Smist, 1980, 1981).

EXPERIMENT I

Experimental Rationale

Experiment I is designed to determine if there are consistent between-group differences in multiple-task performance by comparing the performance of the three strategy groups across a variety of task combinations.

Method

Tasks

Classification. For this task two randomly selected digits between five and eight were presented simultaneously to the subject. The digits varied on two dimensions: size and name. The subject determined the number of dimensions on which the stimuli were alike and then pressed one of three keys on his left-hand keyboard. As soon as the subject pressed a key, the pair was erased and a new pair presented 2 msec later.

Two dependent variables were calculated for each trial: the average interval between correct responses (CRI) and the percentage of correct responses to the total number of responses emitted. The average CRI differs from the more common average reaction time in that incorrect responses are not counted in its calculation. This is, when an incorrect response occurs, the CRI is the time between the preceding correct response and the next correct response including the time during which the incorrect response was made. Both the percentage of correct responses and the CRI were displayed to the subject at the end of each single- and dual-task trial.

Memory. In this task randomly selected digits between one and four were presented sequentially to the subject. The subject retained the most recently displayed digit in memory while responding to the preceding digit. For example, if the first stimulus were "1" and the second "3", the correct response to the "3" would be "1". Responses were made by pressing one of four keys on the right-hand keyboard. The keys were numbered from left to right beginning with "1". The response to the first stimulus of any trial was always "1". As soon as a response was made, the stimulus was erased and the next one presented.

Two dependent variables were recorded: CRI and the percentage of correct responses. At the end of each single- and dual-task trial the CRI and the percentage of correct responses were displayed to the subject. Under dual-task conditions the digits for the classification task were presented on the left side of the display screen; the digit for the memory task was presented on the right side.

Tracking. Two identical one-dimensional compensatory tracking tasks each required the subject to keep a moving circle centered in a horizontal track by making appropriate left-right manipulations of a control stick. One task was controlled by each hand. The inputs to the two displays consisted of the sum of sine waves of .02, .03, .07, .13, .23, .41, .83, 1.51, and 3.07 Hz. The inputs to the two displays were independent. The control systems had mixed first- and second-order dynamics with weightings of 0.10 and 0.90 respectively.

Average absolute error was calculated for each task and presented to the subject at the end of each trial, one indication for single-task trials and two for dual-task trials. The tracking tasks that were controlled by the left and right hands were appropriately offset to the left and right of the display center.

Listening. Each trial in this task consisted of a series of 100 stimulus pairs presented dichotically to the subject. The stimuli consisted of the digits zero through nine and all letters of the alphabet except W, J, H, I, and T. These letters were not used because pretest data indicated that their probability of detection was either very high or very low. The letters were targets and could appear in either ear alone or both ears simultaneously. An equal number of targets was presented to both ears during a trial and could vary from 8 to 12 per ear. Half of the targets during a given trial were paired with targets in the opposite ear, the other half were paired with numbers. The average probability of a target was 0.1 for each ear. If a target appeared in the left ear, the subject pushed his left-hand response key; if it appeared in the right, he pushed his right-hand response key. Stimuli were presented at 56 dB(A) against a background of white noise which raised the overall sound level to 71 dB(A) with a -14 dB signal to noise ratio. The interstimulus interval was 750 msec and the average duration of the stimuli was 325 msec.

Under selective listening conditions subjects concentrated only on one ear, ignoring information presented in the opposite ear. Under dichotic conditions subjects attended to both ears equally. Each pair of stimuli was aligned to give simultaneous onset.

Apparatus

A Processor Technology Microcomputer with a Helios II disc system recorded all responses and performed all timing. This system displayed all inputs for the tracking, classification and memory tasks on a KOYO Model TMC-9M CRT.

Tracking. The forcing function was recorded on a cassette tape and was played on a Phillips Minilog 4 Data Recorder. The output of the recorder was fed through filters that were implemented on two EAI TR-20 analog computers to provide the desired power spectrum of the forcing function. The control system dynamics also were programmed on one of the EAI TR-20 computers. Subjects tracked the function using a pair of Measurement Systems Incorporated Model 541 Two-Axis Gimbal Joysticks. Both sticks were modified to permit movement in the left-right dimension only.

Memory, Classification. Subjects made their responses for these tasks by depressing a key on each of two 4 by 4 matrix-type Micro-Switch Model SW-10196 keyboards mounted in the table in front of them. The keys which were employed had no identifying marks which would connect them to the stimuli.

Listening. The stimuli for this task were created using a VOTRAX Voice Synthesizer at the Naval Aerospace Medical Research Laboratory at Pensacola, Florida. The tape was played on a TEAC A-2340 reel-to-reel tape deck and presented through Koss/E.9 headphones to the subject. White noise was produced using a Grason-Stadler 1724 noise generator and superimposed on the stimuli through a SONY MX-20 sound mixer. Subjects responded to the stimuli by pressing one key on each of the two keyboards described above.

Design

A within-subject repeated measures design with one independent variable, practice, was used. The experiment was conducted on 4 consecutive days. The task presentation order was the same for all subjects.

Subjects

The eleven subjects were right-handed native English speaking males between the ages of 18 and 35 who responded to advertisements placed in the campus newspaper and in campus buildings. They were required to be flight naive with no significant physical handicaps and have good hearing and vision. Subjects were paid $2.50 per hour.

Procedure

Day 1. On this day, subjects performed under single-task conditions only. They received ten tracking trials, beginning with the right hand and alternating between hands on each successive 1-min trial. During the 1-min between-trials rest pause, they received feedback on their performance.

Following the tracking task, the subjects received a total of 12 selective listening trials. Subjects performed six trials per ear, starting with the right and alternating between ears. Each trial lasted 75 sec with a 40-sec rest pause between trials. No feedback was given.

Next the subjects performed six trials each on the memory and classification tasks, alternating between tasks. All trials were 1 min long with a 1-min pause between trials during which the subjects received feedback. The entire first session lasted 1.5 hours.

Days 2, 3, and 4. On the second day of the experiment subjects performed under dual-task conditions. Subjects first performed two single-task tracking trials, one on each hand. The subjects then performed two blocks of five dual-task trials. Each dual block was followed by both a right-hand and a left-hand single-task trial. All trials were 1 min long followed by a 1-min rest pause. Feedback was provided after each trial.

The subjects then performed a pair of selective listening trials, one on each ear, followed by two blocks of seven dichotic trials. Following each dichotic block, subjects performed a pair of selective trials. All trials were 75 sec long, followed by a 40-sec rest pause. At no time during the listening task did subjects receive any feedback.

In the final portion of the session subjects performed two trials each on the memory and classification tasks. Subjects then performed two five-trial blocks under dual-task conditions, each block being followed by a single-task memory and a single-task classification trial. All trials were 1 min long followed by a 1-min pause, during which feedback was given.

Days 3 and 4 followed the same pattern as Day 2. Each session on Days 2, 3, and 4 required approximately 2.3 hours.

Results

Strategy Analysis

The subject's response strategy was classified using the following procedure: First, the percentage of between-task switches was calculated by dividing the number of actual switches by the total number possible. The total possible was equal to 2n-1 where n is the number of responses made to the task responded to the least. For instance, if the subject made 43 responses to the classification task and 23 to the memory task, the total number of possible switches was 2(23)-1=45. Next the percentage of simultaneous responses was calculated by dividing the number of simultaneous responses by the total number possible and multiplying by 100%. For the example given above, the total number of possible responses was 23. Finally, the greatest number of sequential responses to either task was noted.

The identification of the subjects' strategies was straightforward; no arbitrary assignments were necessary. Five subjects used the simultaneous strategy; three, the alternating strategy; and three, the massed strategy.

Single-task Performance

The single-task data from Days 2-4 were analyzed for between-group differences using a two way (group by trials) analysis of variance. No significant main effect of group or group by trials interaction was found for any of the tasks.

Consistent Individual Differences

Before the dual-task data were examined for evidence of consistent between-group differences, the performance on each task combination was

300

examined for differential stability (Jones, 1969). A task (or combination) obtains differential stability when the rank order of subjects remains constant within experimental error from trial to trial. A technique suggested by Bittner (1979) was used to examine each task combination. Stability was obtained on the fourteenth dual-task trial for the memory-classification combination, the eighteenth trial for the tracking-tracking combination, and on the first block of trials (Trials 1-5) for the dichotic listening task (performance was calculated over blocks of five trials on this task combination because the subjects made so few responses).

Multivariate Analysis

The unadjusted dichotic P(A) scores and the scores for the other two task combinations, which were corrected for individual differences in single-task performance, were submitted to a two-way (groups by practice) repeated measures MANOVA. Only two levels of the practice factor, Days 3 and 4, were included in the analysis because neither the tracking-tracking combination nor the memory-classification combination obtained stability on Day 2. The multivariate F for the effect of groups was significant ($F_{6,12} = 4.0860$, $\underline{p} < .02$). Only the univariate test of P(A) was significant ($F_{2,8} = 7.3671$, $\underline{p} < .02$) although the test for the memory-classification combination just missed significance ($F_{2,8} = 3.8964$, $\underline{p} < .067$). The main effect of practice also was significant ($F_{3,6} = 17.7950$, $\underline{p} < .01$). Both the univariate tests on the tracking-tracking combination ($F_{1,8} = 8.4878$, $\underline{p} < .02$) and on the memory-classification combination ($F_{1,8} = 54.6367$, $\underline{p} < .001$) were significant. The group by practice interaction was not significant ($F_{6,12} = 0.6751$, $\underline{p} > .05$).

TABLE I

Dual-task Performance as a Function of
Practice and Response Strategy Group

Group	Tracking-Tracking	Memory-Classification	Listening
Day 3			
Simultaneous	433	5.9	.846
Alternating	332	10.4	.787
Massed	405	10.6	.855
Day 4			
Simultaneous	383	8.2	.850
Alternating	267	13.5	.773
Massed	376	14.7	.862

EXPERIMENT II

Experimental Rationale

The results of Experiment I seem to indicate that the response strategy reflects differences in multiple-, but not single-, task performance. However, it is still possible that the strategy itself is the cause of the pattern of results observed in Table 1: only performance data from Days 3 and 4 were used in the analysis and the response strategy developed on Day 1. Therefore, it is necessary to determine if the significant effect of group indicates between-group differences in information processing or simply reflects the effect of the strategy itself on performance on a variety of task combinations.

One way to determine if different response strategies reflect fundamental differences in information processing is to ask the subject to change strategies. Such a change should result in a minimal disruption in performance if the "natural" strategy is simply selected at random but should disrupt performance severely if the strategy reflects fundamental differences in information processing. However, a large disruption in performance also may occur if the subject attempts to adopt a new strategy after extensive practice with another. Therefore, to prevent an ambiguous result, it is important to provide only enough practice trials to identify the natural response strategy.

Practically, there appeared to be no reason to examine changes from more efficient to less efficient response strategies because there are no real-world situations where the timesharing efficiency of an operator would be degraded deliberately. Thus, subjects always were asked to change to a more efficient response strategy, i.e. from a massed response strategy to a simultaneous response strategy.

Method

Tasks and Apparatus

Only the memory and classification tasks were used. These tasks and their apparatus were identical to those used in Experiment I.

Procedure

Day 1. The subject first performed 12 single-task trials, six on the memory task and six on the classification task. The subject then performed five dual-task trials followed by one trial on each task alone. This testing session required approximately 1.25 hours.

Day 2. The subject performed a total of four single-task trials alternating between the two tasks. The subject subsequently performed five blocks of trials consisting of five dual-task trials followed by one trial on each task alone. This session required approximately 1.75 hours.

Design

A five-group split-plot factorial design was used on Day 2. Subjects were assigned to one of the five groups on the basis of the response strategy used to perform the memory-classification combination on Day 1. All of the subjects using a simultaneous response strategy on Day 1 were asked to continue using that strategy on Day 2 (Simultaneous-Simultaneous) (n=8). The alternating and massed strategy subjects were assigned to their Day 2 groups at random with the restriction that approximately half of each group be assigned to each of the Day 2 strategies. Thus, one-half of the subjects using the alternating strategy were asked to continue that strategy on Day 2 (Alternating-Alternating) (n=3); the other half were asked to use a simultaneous strategy (Alternating-Simultaneous) (n=4). Similarly, about one-half of the massed subjects were asked to change to an alternating strategy (Massed-Alternating) (n=12); the other half, to a simultaneous strategy (Massed-Simultaneous) (n=10).

Subjects

A total of 37 right-handed subjects were recruited through advertisements placed in the student newspaper and in university buildings. All subjects were between the ages of 18 and 35, were native English speakers, and had vision correctable to 20/20. None had any flight training.

All subjects were paid $2.50/hour for participating. The three subjects in each of the five Day 2 groups having the first, second, and third best overall dual-task performance were given bonuses of $10.00, $6.00 and $4.00 respectively.

Results

Strategy Analysis

The strategy analysis was conducted in exactly the same way as for Experiment I. Again, the response strategy used by the subject could be clearly identified. Eight used the simultaneous strategy; seven, the alternating; and 22, the massed.

Day 2 Performance

Because the strategy used to perform the memory-classification combination is known to have major effect on dual-task performance (Damos and Smist, 1980; Damos and Wickens, 1980), it was of no real interest to compare groups using different response strategies. Therefore, all single- and dual-task analyses were conducted within strategy groups.

Single-task performance. Two-way (group by trial) ANOVAs were conducted on the memory data and the classification data for the three groups of subjects who used the simultaneous response strategy on Day 2 and the two groups who used the alternating strategy. The only significant effect was a main effect of trial for the classification task for the simultaneous response strategy groups.

Dual-task performance. To determine if the differences in group performance shown in Figure 1 were significant, a two-way (group by trial) repeated measures ANOVA was performed on the dual-task data averaged over tasks for the simultaneous strategy groups. Both the main effect of group ($F_{2, 19} = 4.5767$, $p < .02$) and trial ($F_{25, 456} = 15.4428$, $p < .001$) were reliable. Additionally, the group by trial interaction was significant ($F_{48, 446} = 1.8069$, $p < .001$).

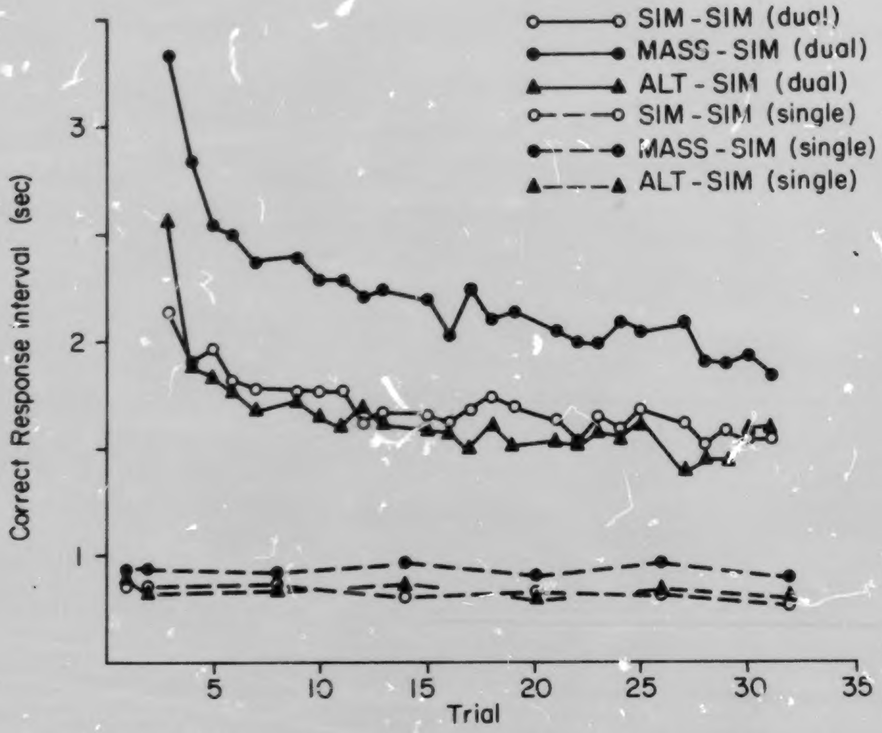


Figure 1. Day 2 single- and dual-task performance for subjects using the simultaneous response strategy.

A similar analysis was conducted on the averaged data for the two groups using an alternating strategy. Only the main effect of trial was statistically significant ($F_{24, 312} = 4.7911$, $p < .001$).

Asympotic performance. The ANOVA conducted on the Day 2 dual-task data for the simultaneous response groups indicates a real performance difference between the Mass-Sim Group and the other two groups (see Figure 1). However, this difference could be transitory, indicating only that massed response subjects have difficulty adopting a simultaneous strategy. It does not imply that massed subjects would not perform as well as the simultaneous and alternating subjects given extensive practice.

One way to determine if the performance difference shown in Figure 1 is transitory is to compare the estimated asymptotic performance for each group. This is accomplished by fitting the data using some predetermined equation and estimating the asymptote from the obtained values of one or more parameters. Initially, the data of a few subjects with very different learning curves were fit using several potentially applicable equations. The equation $CRI=a+b/T$, where T is the trial number, generally provided the best fit across the various learning curves and subsequently was selected for use.

Each subject's data were fit individually and a one-way (group)

ANOVA performed on the estimated asymptotes. The effect of group was significant($F_{2,19}=4.4666$, $p<.03$). The average asymptote for the Massed-Simultaneous Group was 2.002 sec; for the Alternating-Simultaneous Group, 1.504 sec; and 1.620 sec for the Simultaneous-Simultaneous Group.

DISCUSSION

The first experiment described in this paper was designed to determine if the response strategy reflects consistent between-group differences in multi-, but not single-, task performance. The second experiment was designed to determine if the differences observed in the first experiment reflected fundamental differences in information processing or simply the effect of the strategy itself.

Experiment I found evidence of consistent between-group differences in performance across a variety of task combinations. It is interesting to note that none of the groups was superior on all three combinations(see Table 1). Also, the relative performance of the three groups does not seem to be a function of any obvious task characteristic, such as pacing.

Experiment II demonstrated that the Massed Response subjects could not change strategies as easily as the Alternating subjects and, even after practice with the simultaneous strategy, could not perform as well as subjects in the other two simultaneous groups. An extrapolation of the Day 2 performance data from these three groups indicated that the Massed subjects would not reach the same level of asymptotic performance as subjects in the other two groups. Although this extrapolation should be viewed with caution, it and the Day 2 data do seem to indicate that the Massed subjects process information from the memory-classification combination differently from the other two groups of subjects.

It seems then that the response strategy does reflect some individual differences in muli-task information processing. Apparently, the Alternating subjects process information similarly to the Simultaneous Response subjects and can be easily trained to use a simultaneous response strategy. The Massed subjects, on the other hand, appear to process multi-task information basically differently from the other two groups of subjects and seem to have difficulty changing strategies. Thus, it may be necessary to select individuals for particular combinations and then train them to perform the combination in the most effective manner possible.

REFERENCES

Bittner, A. C., Jr. Statistical tests for differential stability. Proceedings of the twenty-third annual meeting of the Human Factors Society. Santa Monica, CA.: Human Factors Society, October 1979.

Damos, D. L. and Smist, T. E. Individual differences in dual-task performance. Amherst, N.Y.: Dept. of Induatrial Engineering, State University of New York at Buffalo, June 1980.

Damos, D. L. and Wickens, C. D. The identification and transfer of timesharing skills. *Acta Psychologica*, 1980, *46*, 15-39.

Jones, M. B. Differential processes in acquisition. In E. A. Bilodeau and I. Mc D. Bilodeau (Eds.), *Principles of Skill Acquisition*. New York: Academic Press, 1969.

SIMULATION TECHNOLOGY AND THE FIXATION PHASE

Edward A. Stark
The Singer Company
Link Flight Simulation Division

ABSTRACT

In 1962 Fitts proposed a 3-phase model of skill learning. The model was derived in large part from the analysis of flight control tasks, and from interviews with physical education instructors. In subsequent research insights were developed into the fundamental nature of the learning functions taking place during second, or fixation phase of skill acquisition. During the same time period increasing use was being made of flight simulators to provide training which could not practically be given in actual flight.

Simulator technology advanced rapidly as engineers developed better ways of reproducing the crew's cockpit, instrument, aural, motion and visual environments. Major simulator users became convinced that flight-like practice in flight-like environments would insure the efficient development of capabilities for skilled performance.

Continuing advances in simulator technology have produced almost totally real instrument and aural cue systems, and a number of specialized g-cuing systems are rapidly being matched to the sensory capabilities and needs of flight crews. Intensive research in visual cuing is providing more and more realistic visual scenes and in general, flight simulators are becoming more real, more complex and more expensive.

It is suggested that the fixation phase does not require a flight environment or even a reproduction of the flight environment for effective learning and transfer. While similarity between training and operational environments appears to be essential for transfer of training to take place, the concept of similarity may not include the total complex of stimuli available in flight as perceived by the instructor or the simulator designer, but may include only the stimuli perceived by the student as he observes, practices and validates specific tasks within the period of response fixation. It is suggested that special attention be given to the task information required and the learning functions involved in the fixation phase of skill learning.

1. INTRODUCTION.

Skill training has almost always relied on some sort of simulation from the use of wooden swords and blunt arrows through training aircraft and various electronic devices. The transfer of skills from the simulator to the operational setting has always been assumed to require some degree of similarity between the two environments; hence, the term "simulation". Oddly enough, attempts at defining "similarity" as it relates to the processes by which complex skills are learned have been few and far between. Similarity is generally taken to refer to identifiable elements of the operational environment, incorporated in the training setting with little consideration for the ways in which these elements are used by the skilled operator or perceived by the trainee.

Recent rapid increases in the use (and utility) of sophisticated simulators in skill training have resulted in, and to a large extent from, brute force applications of technological advances in increasing the objective similarity of training and operational settings. These applications of technology have increased simulator realism in an engineering sense, but have accounted too little for the ability of the student to perceive and employ operational reality in skill development. As a result, trainees have been able to practice in more and more realistic task settings but there have been few improvements in the guidance provided them in the identification and interpretation of the information represented in the simulator in support of practice. Simulators have been used to remarkable advantage by providing relevant, dynamic and important task information, but neither maximum training efficiency nor maximum training effectiveness will ever be achieved until this information is organized not only for task relevance but also for relevance to the processes by which learning occurs and skills are developed.

The traditional use of textbooks, lectures, pictures, part-task trainers and other essentially "non-real" settings in flight training programs reflects a recognition of the general idea that complex skills are not always learned most efficiently as complete entities. Unfortunately, it also reflects a training system philosophy directed more at the convenient administration of training than at the management of learning processes. In general, phased training settings are defined and organized to support cognitive learning and perceptual-motor practice in mission-derived training tasks with little formal consideration of the variety of psychological functions involved in the development of total task proficiency.

When Fitts proposed his model of skill learning processes, he suggested that each skill consists of cognitive, perceptual and psychomotor elements. The cognitive element is concerned with the knowledge required to describe the task to be performed. The cognitive element includes the words used to describe the task and the procedures, techniques, and standards which define its essential characteristics. The perceptual element includes those processes by which task information is recognized, discriminated, interpreted and used in controlling task performance. The psychomotor element includes the perceptual and neuromuscular activities which constitute task performance. Cognitive elements are usually learned early in training to form a basis for understanding the task, for implicit manipulation of task elements and conditions, and for organizing practice in the psychomotor activities which constitute task performance. Cognitive task elements are generally considered to account for a small proportion of the overall skill and for a similarly small proportion of the time and effort spent in skill development.

It is quite likely that cognitive learning never really ends, but continues as the learner becomes more and more familiar with the task and with his own capabilities in performing it. At the same time, the bulk of cognitive learning appears to take place prior to the beginning of the next phase of learning which Fitts called the fixation phase.

2. THE FIXATION PHASE.

In the fixation phase of skill learning several different processes take place as relatively discrete patterns of implicit and explicit behavior are practiced and perfected. A repertoire of perceptual and psychomotor skill elements is developed. Increasing levels of skill are reflected generally in increasingly rapid and accurate overt performance. The fixation phase and the next phase - the automation phase - overlap significantly, but the fixation phase is characterized by the development of individual but interrelated perceptual and psychomotor skills. In the automation phase these individual skills are integrated into more complex performance capabilities and the individual skill patterns become autonomous requiring little, if any, conscious attention on the part of the operator once they have been selected and initiated.

The psychological events taking place during the fixation phase can require weeks, months or even years of practice. Fitts cited data collected by Snoddy in 1926 to illustrate the time required in the development of a simple but unique skill. Snoddy's subjects practiced tracing within the narrow borders of a star pattern. The task was made novel by requiring the subjects to observe the star, the pencil and the hand holding it in a mirror. The subjects practiced each day for 59 days, continuing to show improvement in time and error scores throughout the entire 59 day period. Fitts also quoted data collected by Crossman who measured the performance of cigar makers. Crossman's subjects continued to show improved performance over a period of ten years.

Obviously the performance of complex skills can be improved with practice over long periods of time It is almist as obvious that most formal flight training programs cannot provide enough practice for many pilot skills to reach anything like an asymptote. While many skills are undoubtedly perfected in line flying operations, many are neglected beyond the stage of minimal proficiency, for lack of adequate practice settings, resources, facilities and time. Unfortunately, many skills required of combat pilots especially, are never really practiced at all except in combat, for reasons of safety and economy. While simulators are being used for training some air combat skills, some skills are being neglected because of the tacit acceptance of the concept of total simulator realism as the only valid, feasible practice setting for learning some complex skills.

Total flight simulator realism is beyond the capabilities of the current technology and it will continue to be for the foreseeable future. In some information areas nearly total realism has already been achieved, but the highest possible overall fidelity in the representation of vehicle motion and in the simulation of out-of-the-window visual scenes falls far short of reality and will continue to fall short for a long time in the future. While this would seem to severely limit the amount of contact flight training possible in simulators, it could, in fact, lead to a new and higher level of reliance on simulators for training. Skill learning does not and never has required complete physical correspondence between training and operational settings. Somehow the process of generalization permits transfer from one setting to another even when quite noticeable differences exist between them. In addition, Fitts' model makes it quite clear that efficient perceptual and psychomotor learning involves something other than simple repetitive practice in "real" task environments. The fixation phase of the model is a recognition of fundamental differences in the processes by which the various parts of the skill repertoire are developed and of the need to address these processes specifically.

Closer attention to the learning processes and to the ways in which task information relates to them can lead to more effective use of the available simulation technology. Relatively discrete practice settings might be developed to support more complete learning of discrete but related skills which could then be integrated in the automation phase of learning in a sophisticated simulator or in the prime system itself. This approach to training could increase the efficiency of skill learning; but, equally important, it could place required attention on a variety of perceptual and psychomotor skills which are currently neglected or only superficially developed. Limitations in the potential for realism in visual simulation systems, in particular, is currently focusing attention on the minimum information needed for training; this in turn will lead to a more thorough analysis of, and insight into the processes by which task information relates to training. Finally, training administrators, trainers, and training device designers will begin to look back at Fitts' fixation phase as a learning arena in which repertoires of essential skills can be developed in less than totally real training settings.

Fitts' postulation of a fixation phase in skill learning was a recognition of the fact that the trainee's perceptions and his approach to learning change as he develops a repertoire of essentially discrete task-related skills. Much of the work in the Aviation Psychology Laboratory in the late 40's and 50's was devoted to establishing the ways in which psychomotor skills are developed. Recent work in the Air Force's Human Resources Laboratory, in cognitive pretraining and backward chaining, and the work of Roscoe and his students at the

University of Illinois have expanded our knowledge of the perceptual and learning processes, while at the same time contributing to more effective learning in a variety of on-going training programs. Also, work in the Pennsylvania State University's College of Health, Physical Education and Recreation is providing additional information about some of the processes which take place in the fixation and automation stages. Christina and others have found that skilled performers do a great deal of anticipating; that is, they use specific cues and cue patterns not only to guide specific responses, but they also use task-related information to develop event time histories which can be used in anticipating future output requirements. Anticipation is essential under conditions when time constraints and, perhaps, threshold problems do not permit classic stimulus-response relationships to control operator outputs. This type of information adds to our understanding of the processes by which complex skills and skill elements are mastered and entered into the student's repertoire in the fixation phase. Incorporating this information in flight training programs will require additional research and additional insights into the way students approach learning situations. In the meantime, it is important to think about the basic elements of which flight skills are composed and of the ways in which these component skills can be most efficiently and economically developed.

3. SIMULATOR APPLICATION.

Most maneuvers can be described as unitary events involving sequences of control inputs, and defined by systematic changes in attitude, airspeed, flight path and aerodynamic status. Unfortunately, most maneuvers are sufficiently complex that the relationships among flight conditions, control inputs and aircraft responses remain a total mystery for a relatively long and frustrating period of time for the pilot trainee. When the aircraft was the primary training medium, learning involved flight manuals, texts, lectures and a lot of practice in sorting out and responding to a relatively complex and confusing practice environment. The first Link trainer and a variety of other ground training devices were developed to reduce the cost and complexity of training. They also isolated important task information, permitting a clearer perception of relevant control dynamics and a simpler and less ambiguous practice environment.

As the value of ground trainers became more and more evident, attention shifted from supporting specific flight task learning processes to providing more and more of the stimuli associated with actual flight. Instead of providing simple, high-fidelity settings for the development of individual skill proficiencies, simulator technology tended to swamp the student with practice settings as complex and as confusing as the flight environment itself. Early trainers provided relatively simple practice settings not so much by intent, but because of inherent limitations in the available technology. Today, task-relevant information can be stored, processed and displayed with rapidly increasing efficiency. But these capabilities are typically used for increasing overall realism rather than in support of simple but crucial training objectives.

It is suggested that today's advanced simulation technology be applied in the support of individual training problems within the context of the fixation phase of learning. This means that specific difficult and important skills and skill components should be trained outside the whole task context in low-cost but high-fidelity training settings designed to mediate only that information relevant to a specific task or task component.

This approach is intended to accomplish three primary objectives: First, to assure that many otherwise neglected skills receive the time and attention they require for effective flight crew performance; second, to provide essential skill training with minimum use of complex simulation and, finally and perhaps most important, to provide unambiguous practice settings which require the student to deal, at least in the early stages of training, with only the information relevant to the specific training objective at hand. The identification of the psychological processes involved in the learning of typical flight control skills may help to identify some of the simpler settings which might be developed for more economical training. More important, it could help to develop settings for more complete learning of skills and skill elements which are currently neglected both in flight and in simulator training.

4. LEARNING FUNCTIONS.

4.1 The Loop. The loop is a simple enough maneuver as viewed from the ground, from the point of view of the skilled pilot, or as seen by the engineer whose device is expected to simulate it. From the point of view of the novice pilot, however, it is a disturbing sequence of rapidly changing conditions over which he seems to have little obvious control, and which frequently turns into a stall or a high-speed rolling dive toward the ground. From the point of view of the learning psychologist, the loop involves an appreciable number and variety of sensations, perceptions, interpretations, anticipations and control sequences and modulations. At least four psychological functions must be exercised in developing proficiency in executing the loop, in addition to the more obvious cognitive components of the skill.

4.1.1 Procedural Functions. Traditionally, procedures are considered to be discrete control events executed in a fixed sequence within a relatively rigid time frame. Procedural skills are usually learned in relatively simple settings, by rote. They are usually readily verbalized and involve a somewhat unique transition from cognitive to motor learning since they involve both an intellectual element and the execution of simple muscular actions. While most complex flight maneuvers are usually not considered procedural, it is useful to think of these tasks in a procedural framework because it is useful in guiding the learning of the more difficult perceptual and psychomotor activities, it is also a reminder that seemingly complex skill elements might also be learned at least in part in relatively simple settings.

Flying an acceptable, aerodynamically consistent loop includes a sequence of more or less pre-defined control inputs. Understanding the sequence of inputs and being able to execute them in order is an important part of the skill. Typically, the loop begins with a specific power setting and airspeed, and an increase in back pressure at a prescribed rate. Eventually, the back pressure is released, reapplied, and finally released again at the end of the maneuver. In addition, power is adjusted throughout the maneuver in a similarly procedural fashion.

If the aircraft is propeller-driven an additional procedural element must be mastered: as airspeed and power settings change, the pilot applies rudder pressure to maintain a constant heading and to keep the wings level. These control inputs are procedural in the sense that they can be memorized as a part of the task requirement and initiated more or less by rote. Sometimes novice performance is recognized by the "mechanical" approach used in performing maneuvers. The pilot knows the inputs he must make; he knows the general sequence in which they must be made; but he has not yet learned to recognize and interact with the information reflecting inaccurate control inputs or changes in the flight environment. This tends to illustrate that while knowing and performing the procedural task elements provides a systematic framework for effective practice, the more complex perceptual and motor elements must also be learned in attaining complete skill proficiency.

4.1.2. Perceptual Functions.The loop requires that the student learn to perceive and interpret a number of complex patterns of task-relevant information. A variety of sensory systems is involved, with primary emphasis on the interpretation of information from the visual scene. Information from the various inertial sensors is also important because it is generally perceived in advance of correlated visual events, once the pilot begins to recognize and correlate the information from the various sensory systems. A major perception to be learned in the development of skill in executing the loop is in recognizing the rate of pitch change required to maintain positive pressure in the seat throughout the inverted phase of the maneuver, without inducing a stall or a departure. Rate of pitch change is sensed early in the loop by observing the motion of the nose with respect to the horizon, but changes in control forces and g-loads and the sounds of the engine and the airflow over the cockpit also contribute to the perception of pitch rate. As the loop progresses, non-visual cues become increasingly important as visual contact with horizon is lost, unless the pilot chooses to use instrument cues to supplement those available outside the cockpit.

The perception of changes in aircraft heading is also important, and like the perception of pitch rate, becomes more difficult as the aircraft reaches the inverted position. Inertial cues are important here also as they accompany changes in roll angle or yaw which could result in changes in heading. Even though torque corrections are made on a quasiprocedural basis as the maneuver progresses, accurate rudder applications require immediate visual and inertial information as the maneuver is developed.

Roll attitude is perceived early in the loop by reference to the horizon; but, again, inertial cues are important in recognizing and correcting deviations from the wings-level condition throughout the maneuver.

Loops are usually practiced at altitudes chosen to provide a substantial margin for error, but the transfer of skills learned in aerobatic practice is most crucial when they are needed in low altitude flight. Commercial pilots rarely enter loops at low altitudes by choice, but require skill in controlling the aircraft at all altitudes and in all attitudes as the aircraft's performance is influenced by inadvertent control inputs, turbulence and weather. Air combat pilots also require skill in aerobatic flight (and in loops) at all altitudes and so a need exists for skill in another perceptual task element. Prior to beginning the loop, the pilot must recognize the space available to him in performing it, not so much as a number of feet of altitude, but as an estimate of available-versus-required vertical separation from the ground. Different loops require different amounts of space with good, well-executed loops requiring less space than poorly-executed loops; with practice, the pilot learns the amount of space required for his own worst-case performance.

The perception of altitude is important at the beginning of the loop and at the end; but the perception of altitude rate at the end of the loop is also a crucial pilot function. Again, altitude rate in feet per second is relatively meaningless to the pilot: for the pilot, altitude rate at the end of a loop must be perceived in relation to the known characteristics of the aircraft. These become known as the pilot is exposed to the information provided by the visual scene, the instruments, the sounds and the control and inertial forces accompanying control inputs under various circumstances and conditions.

The fixation of perceptions associated with normal maneuvers is insufficient in preparing the pilot for the integraton and refinement of more complex skills. Flight manuevers are not always executed as described in the training manual, due to variabilities in human performance and unanticipated or misperceived conditions in the flight environment. As a result, the training setting must be designed to provide information associated with normal performance and also that information, sometimes only subtlely different, which defines conditions which are not normally encountered. Frequently these kinds of information are crucial in differentiating among situations making drastically different demands on the pilot. In the loop as in other maneuvers it is important that the consequences of incorrect control inputs be represented to permit the pilot to perceive and eventually anticipate and avoid the consequences of improper inputs.

Perception and perceptual flight task elements are generally of interest only as they relate to the control outputs and modulations required to operate the aircraft and its systems. Once important perceptions begin to be developed they must be associated with relevant control actions so that essential perceptual-motor skills can be developed.

4.1.3 Perceptual-Motor Functions. Procedures are well established control activities accomplished in a fixed sequence; they are defined as "well established" because they do not involve any unique patterns of neuromuscular response. They include actuating a switch, turning a knob, or moving a control in one direction or another. Procedural learning does not develop these muscular patterns, but simply attaches them to specific unambiguous bits of information. Perceptual-motor learning not only attaches specific control outputs to both discrete and complex information patterns, but it also develops skill in modulating control inputs based on changes in stimulus patterns. In the flight environment patterns of task relevant information frequently change too rapidly for the pilot to recognize, interpret and respond to the change even though a change in control output is required. Under these circumstances perceptual motor activity includes both the modulation of control outputs as relevant information is observed and interpreted and the development of skill in anticipating control requirements.

In the loop the primary perceptual-motor element is in controlling pitch rate through the application of back pressure. The pilot maintains a rate of turn in the vertical plane with elevator inputs which are initially procedural. As he learns to interpret the information accompanying the maneuver he begins to modify inputs to the elevators to produce changes in the various stimuli reflecting aircraft performance. As learning progresses the pilot develops the ability to make more and more accurate control inputs and modifications while at the same time learning to perceive patterns of performance unique to the specific aircraft.

Other perceptual motor responses are also learned in the loop. Control of roll attitude is highly dependent on the perception of both visual, inertial and, in many aircraft, control force cues. The initially procedural torque correction is also refined as the cues to yaw and roll accompanying changes in power setting and airspeed begin to be recognized. Power settings also begin as procedural inputs, but as the pilot recognizes discrepancies between indications of pitch rate, g-load and the sound of the engine and the air flow and those experienced in "good" loops and in demonstrations, power control inputs are modified to reduce and avoid these discrepancies.

4.1.4 Anticipation. There is always some delay between the recognition of a control requirement and the initiation of the control input. There is also a delay between the initiation of input and the change in performance it is intended to effect. A major element of perceptual motor learning is in accounting for these delays through the development of timing skill and through the development of skill in anticipating control requirements at some time in the future. Control system delays and dynamics require timing skill while the rapidity with which many flight conditions change requires the pilot to learn to predict events from a recognition of trends in the performance of the aircraft or in the changes taking place in the environment.

Anticipatory behavior can be learned as a part of the perceptual-motor learning processes taking place in actual flight or in a simulator having the ability to represent relevant task information and task dynamics. It may be possible also to learn anticipatory behavior in simpler training settings if these settings can represent the cues and the dynamics essential to the anticipatory behavior itself.

4.1.5 Judgment. Pilots operating in flight environments begin to learn to exercise judgment as soon as they begin to exercise control of the aircraft. Judgment is concerned less with the quality of specific control inputs than with the recognition of control requirements and their application for maximum safety and performance efficiency. Traditionally, judgment is taught as a cognitive skill with heavy reliance on the observation of skilled performance in day-to-day flight and flight training situations. Pilots, whether they are experienced or novices constantly exercise judgment, not only in the selection and execution of a flight path, but also in the operation of the aircraft and its systems. Judgment is exercised by assessing existing conditions and responsibilities and selecting a control strategy designed to maximize safety, economy, and chances of mission success.

While judgmental training is probably not generally considered to be a major function of the fixation phase of skill learning, the fixation phase, particularly if it occurs in the simulator or in a series of simpler training settings, can be crucial to the development of effective judgmental skills.

Simulators and other non-real training settings are frequently chosen for accomplishing flight training objectives because, beside being economical, they are safe. A simulated crash or mid-air collision poses no hazard to life or property, permitting trainees to explore and become proficient in situations which could not be explored in actual flight. In the loop, for example, the pilot can learn to estimate the vertical space available to him at low altitude and to experience the dynamics of the loop in environments he would not expose himself to in flight. As an air combat pilot he can also learn to modify the basic loop to make an attack or to evade one, at the same time learning to sense the energy available to him throughout the process. Obviously, simulation technology has a great deal to offer in the development of important skills, but it also has the ability to shield the student during a major part of his formal trainee experience from the need to make many kinds of judgments which he would be required to make if training were given in the aircraft itself. Consideration must be given to the judgmental aspects of each flight task to insure the development of training settings capable of supporting them as well as the perceptual and perceptual motor skill elements.

The loop is a complex maneuver which can eventually be exercised almost as a procedural task; as it becomes fixed in the pilot's skill repertoire it can be accomplished with little conscious thought or attention if it is always performed in the same kind of flight environment. Other difficult maneuvers are not so readily automated because they are performed in a variety of environments which force the pilot to develop a number of perceptions and discriminations all relevant in the performance of the maneuver, but which reflect significantly different task conditions. Here the characteristics of the settings in which perceptual, psychomotor and judgmental skills are to be learned are defined more by the conditions under which the maneuver is performed than by the dynamics of the maneuver itself.

4.2 The Rapid Deceleration. The helicopter pilot's rapid deceleration is a good example of the kind of task in which the task environment makes unique demands on the fixation phase of learning. Rapid decelerations are practiced in many kinds of terrain and involve bringing the helicopter to a stop in a position providing concealment or protection from hostile observation or from engagement by air defense weapons. The maneuver must be performed quickly: vulnerability is proportional to the time required to disappear behind an obstacle. Because the maneuver must be done quickly there is always the possiblity of selecting the wrong place to hide, overstressing the helicopter and striking the ground or an obstacle.

4.2.1 Procedural Functions. The rapid deceleration maneuver is procedural in the sense that it requires specific sequential inputs to the collective, cyclic and antitorque controls. The procedure is similar to that used in an autorotation and is not really learned as a unique part of the pilot's repertoire of evasive actions. The perceptual and perceptual-motor functions are also quite similar to skills learned earlier, but involve a significant degree of concern for tactical as well as purely psychophysical responses.

4.2.2 Perceptual Functions. A major perception in low altitude tactical helicopter operations concerns the pilot's awareness of the characteristics of the terrain over or through which he is flying. Since he expects to be observed and engaged from the ground he must learn to perceive line-of-sight relationships between his position and potential threat implacements and between these areas of potential threat activity and other areas on the ground which might conceal him from observation. His perception of the terrain is strongly influenced by the information provided in the form of maps, photographs and briefings, but a significant amount of perceptual skill involves observing relationships in the terrain, as they developed, in real time. The location of features capable of providing covering concealment for threats and for friendly elements and for the helicopter itself must be recognized. In addition, the pilot must recognize possible routes of movement for surface units and systems, possible fields of fire for both hostile and friendly forces, and the landmarks, if any, identified in briefings and in pre-mission planning. He must also be able to recognize movement and indications of movement in the terrain and indications of friendly and hostile activity. The pilot must be able to discriminate among different kinds of activity identifying specific weapon types, locations and trajectories in order to plan and execute evasive maneuvers.

The perception of the tactical environment is crucial in permitting the pilot to accomplish his mission; but equally important, he must learn to perceive the characteristics of his aircraft. Specific perceptions must lead to specific relevant and accurate system performance. In the rapid deceleration the pilot must be able to anticipate the rate of descent of the helicopter and the lead time required to bring it to a hover in close proximity to the ground. This means that he must perceive vertical (or slant) velocity in terms of time-to-go and he must perceive the time and control inputs required to establish, maintain and terminate the descent. Since the rapid deceleration will frequently be used to place the helicopter behind an obstacle, the pilot must also perceive his rate of closure with the obstacle and the capabilities of the helicopter for achieving and terminating motion around it.

Most of the perceptions discussed so far are primarily visual. The pilot observes a variety of visual scene elements, and through practice learns to organize these elements into patterns of shapes, shades, colors and movements having special meaning for his control responsibilities and mission functions. Many perceptions of helicopter capabilities also involve the non-visual sensory systems including the receptors responding to control pressures, skeletal movements, vibrations, seat pressures and gravitational forces. While visual information is of primary importance, the generally shorter response latencies of the non-visual sensors make them important in developing maximum control proficiency and in helping the pilot to avoid overstressing the helicopter while maximizing the maneuvering capabilities available to him.

Since military helicopters operate in a wide variety of terrain, tactical, lighting, weather, load and visibility conditions, the pilot must develop perceptions peculiar to these differences. The ability to recognize and discriminate among elements and features in the terrain changes drastically with viewing conditions and, in effect, requires the development of a large number of related but distinctly different perceptions.

4.2.3 Perceptual-Motor Functions. The best rapid acceleration is the one which exposes the helicopter to ground observation for the least amount of time and which does not result in damage to the helicopter and its systems. The pilot must learn to initiate the maneuver without inducing a negative g-load on the helicopter and he must terminate it in a hover as near to the ground and as close to the source of concealment as possible. This means that each perception must be associated with a related set or sequence of control inputs. Since each rapid deceleration occurs in a different perceptual environment, many different perceptual-motor responses must be developed. It is likely also that some sort of generalization still must be developed to account for small, but crucial differences in maneuver conditions.

4.2.4 Anticipation. Being "ahead of the airplane" is a crucial part of every pilot's job. Anticipation is an even more important part of the tactical pilot's responsibility since he must anticipate not only what his own aircraft will be doing in the next few seconds, but he must also be able to predict, to the extent possible, the actions of many different elements in the tactical environment. In this context the anticipatory function includes a significant element of risk-taking as the pilot attempts to predict events to which he will have to respond, while at the same time estimating the probability of their occurrence.

4.2.5 Judgment. Anticipating and estimating the probability of occurrence of important events is the essence of judgment. Judgmental skills are developed through exposure to circumstances representing various levels of ambiguity and criticality and through the observation of the consequences of the decisions made with respect to them. Judgmental learning is also a function of the pilot's confidence in his own ability to correctly interpret complex situations and to interact effectively with them. The helicopter pilot, faced with the possible necessity of performing a rapid deceleration, must be confident that his estimate of the situation is valid and that his ability to respond to it is likely to be effective. It would appear that simulation and other even simpler synthetic training environments could contribute heavily to training in the processes by which effective judgments are made, provided simple statements of training objectives and definitions of relevant task information can be developed.

Like the pilot learning to perform the loop, the pilot learning to use the rapid deceleration as a tactical maneuver is faced with requirements for many different kinds of learning. Since his task is a flight control event as well as a tactical maneuver the skills to be developed are more complex than those required in the aerobatic loop, but they are directly related to specific kinds of definable information and they require the exercise of learning processes which can also be identified for efficient and, probably more economical training. Neither the loop nor the rapid deceleration requires a single complex training setting, but both require careful analysis of the skills and processes involved in their practice and mastery.

The performance of many flight tasks is influenced not only by the characteristics of the aircraft and the environment in which it is flown, but also by the performance of other dynamic elements in the flight environment. The fighter pilot flying an air combat mission, for example, must be proficient in performing a wide variety of flight tasks and maneuvers. He must also be proficient in adjusting his performance to the performance of other friendly and hostile elements on the ground and in the air, frequently in an extremely short period of time. Here again, the skills involved in air-to-air combat are complex and the levels of proficiency required for mission success are very high. For practical reasons it is quite likely that very few are ever developed to anything like the highest reasonable levels in conventional flight training programs. An analysis of a typical maneuver may help to focus attention on skill elements which could be developed in practical non-combat, non-flight training settings.

4.3 High Speed Yo-Yo.

A pilot overtaking a hostile aircraft attempts to place the aircraft within the range and launch limits of his weapons. The target aircraft, at the same time, may attempt to evade by turning at a greater rate than his attacker. If the attacker's velocity is greater than the opponent's the centrifugal force of the corresponding turning maneuver places him outside the turning plane of the target and thus places the target outside the effective zone of the attacker's weapons. Thus, the attacker must lose velocity while maintaining visual contact with the target, and before the target has time to initiate a different evasive maneuver. The high-speed yo-yo accomplishes this, and is initiated by lifting the nose with top rudder to bleed off energy, maintaining visual contact with the target throughout the maneuver. As velocity begins to match that of the defender the nose is lowered again to establish a pursuit relationship.

4.3.1 Procedural Functions.

The two primary elements of the maneuver are the visual interpretation of target/attacker relationships and relative velocities and sensing and controlling the available energy level throughout the maneuver to permit the pilot to place the defender in his weapon envelope. Once the attacking pilot senses an overshoot or speed advantage he uses top rudder to raise the nose, sensing the loss of energy through changes in g-load, angle-of-attack and consequent changes in aerodynamic buffet. As energy is dissipated bottom rudder is used to lower the nose, to enter the turning plane of the opponent. The task is procedural in the sense that a prescribed sequence of control inputs is made, but each input's characteristics are determined both by the nominal demands of the maneuver and by their effects on the velocity and flight path of the aircraft. They also depend heavily on the behavior of the target as the procedure is executed. Typically, the yo-yo is practiced initially without a target, permiting the pilot to exercise procedural inputs and, most important, to learn to recognize the cues which tell him what his aircraft is doing and how much maneuvering energy is available throughout the maneuver.

4.3.2 Perceptual Functions.

Three perceptions are of extreme importance in the high-speed yo-yo. Two are largely visual, but are probably intimately involved with the pilot's ability to project the performance of his own aircraft (a multi-sensory perception) on the performance of the defender. The first perception concerns the relative velocities of the two aircraft. The initiation of the high-speed yo-yo is predicated on a positive rate of closure with the pursuing aircraft gaining significantly on the defender. Depending on the size of the opponent and the viewing conditions, the fighter pilot may sense significant rates of closure with another fighter at distances as near as four or five miles. If the defender initiates a defensive turn, the difference in velocity can make it impossible for the attacker to stay in the defender's turning plane and, more important, to bring his weapons to bear. As a result it is important for the attacker, first, to sense the fact that he is overtaking and, second, to accurately gauge the rate of overtake and its effects on his tracking solution. Estimating the rates of overtake also tends to dictate the point at which the yo-yo must be started and the degree to which the nose must be elevated.

The third perception concerns the performance of the aircraft itself, both in the normal training environment and in the conditions typical of combat operations. These latter perceptions are critical in the development of combat proficiency but are generally neglected in conventional training because they involve unacceptable hazards and inordinate expense. For these reasons, they require careful attention in the development of safe, economical and effective training settings.

Aircraft performance is described in a number of different parameters including airspeed, sink rate, climb angle, attitude and turn rate; but for the fighter pilot the aircraft's energy level during an engagement is of utmost importance. If he has energy available he can use it to increase velocity in attacking or evading an opponent or he can use it to gain altitude to achieve an advantage. If the pilot has altitude he can use that to provide energy, if he can avoid getting too close to the ground. The energy level of the aircraft is sensed through control forces, accelerations on the pilot's body, sounds, changes in the visual scene, and in the cockpit instruments, when the pilot can spare the time and attention required to read them.

The perception of energy level requires the interpretation of a variety of information throughout the flight envelope of the aircraft, sensing conditions along the edges of the envelope, that is, at the highest angles of attack, the lowest airspeeds and altitudes and the greatest g-loads at which the aircraft (and the pilot) can operate effectively. These perceptions are hard to develop because of the hazards involved and because of the effects of changing flight environments and aircraft configurations, but they are crucial in permitting the pilot to maintain positive and effective control when flying against a similarly proficient opponent.

Most of the perceptions required in air-to-air combat operations are difficult to develop because of the natural hazards of the environment; but, perhaps more important, because of the amount of information involved and the number of sensory systems required to develop unique and unambiguous perceptions. In addition, of course, the air combat environment can change more rapidly than the sensory systems can adjust, requiring a high level of skill in anticipating and predicting future events requiring new or modified responses.

4.3.3 Perceptual Motor Functions.

The perceptual-motor responses required in executing the high-speed yo-yo are not particularly difficult, but they cannot be executed by rote except in highly artificial practice environments. In operational flying and even in conventional in-flight practice the rate at which rudder inputs are made, the amount of roll-out required, the height of the yo-yo and changes in g-load and angle-of-attack all depend on the effects of each input on the next and, when a target aircraft is involved, on the actual closure rate and on the behavior of the opponent throughout the maneuver. In operational flying it is unlikely that a high-speed yo-yo would ever be completed as the opponent senses and counters it at sometime during the maneuver. The perceptual-motor activities used in the yo-yo to optimize tracking and energy conservation are not unique to the yo-yo, but are probably adaptations of a repertoire of previously learned responses. Early flight training and the basic flight maneuvering syllabus are designed specifically to help the pilot to learn the fundamental performance characteristics of the aircraft.

4.3.4 Anticipation.

During air combat maneuvering practice the pilot learns to perceive the flight path of the opponent at extreme range and to recognize target attitudes and aspects associated with various flight paths. At the same time, he learns to anticipate flight path and velocity changes both by understanding air combat tactics and their relationship to various flight and tactical environments and by recognizing changes in attitude before they result in actual changes in flight path. Some of the pilot's ability to anticipate his opponent's behavior results from his own experience as a defender; but much of it comes in conventional training from familiarization with the opponent's aircraft and his tactics. In addition, it comes from as much practice as possible in flying against aircraft tactics similar to those used by the expected opponent.

4.3.5 Judgment.

The development of judgmental skills involves the development of skill in estimating risks associated with various possible courses of action. The development of judgmental skills in air combat training, as in training in other flight tasks, can suffer from increased exposure to synthetic training settings at the expense of training time spent in actual flight. Unless specific provisions are made for the support of judgmental training objectives. Very little judgmental skill is involved in practicing the yo-yo other than that involved in flying the aircraft in proximity to the ground, clouds and other aircraft. At the same time, the pilot must eventually develop judgment not so much in executing the yo-yo as in applying it in a variety of tactical situations.

5. SUMMARY AND CONCLUSIONS.

Simulator technology has made remarkable advances over the past 30 or so years as the capabilities of data storage, processing and display systems have been improved. Much of the technology has been applied in the development of complex and increasingly "real" training settings. More attention is needed in the development of limited but high fidelity training environments for the training of important and difficult tasks and task elements. Unfortunately, there is a tendency to treat specific tasks as unitary functions and to treat task learning in terms of practice in the actions observed in skilled task performance. In reality, most flight tasks consist of a variety of subtasks and task elements whose nature and interrelations change as task proficiency develops. Also, most flight tasks involve more than one psychological function, each facilitated by a somewhat different kind of task-relevant information. More careful analysis of flight tasks and related learning processes is needed to permit the more relevant application of the available technology in training support, within training settings designed not so much for "realism" as for the support of the fixation phase of skill learning as defined by Paul Fitts.

More careful application of the technology available to support training can improve training efficiency and economy; it can also support training in skill areas which have heretofore been neglected. Many important skills, especially those involving flight control and marginal conditions, and those relating to air combat have been neglected because of the assumption that they require complex and expensive training settings and sometimes because of the assumption that the required simulation is beyond the capabilities of the technology. Analysis of the psychological events and learning processes associated with these skills can lead to the development of simple, inexpensive and effective training settings. Complex simulators will always be required in training some complex and difficult skills and in integrating others. It is unlikely, however, that they are essential in training many individual skills and skill elements of importance.

REFERENCES

Christina, R. W., Proprioception As A Basis Of Anticipatory Timing Behavior, in Stelmack, G. E. (Ed), Motor Control, Issues And Trends, Academic Press, New York, 1976.

Fitts, P. M., Perceptual-Motor Skill Learning, in Melton, A. W. (Ed) Categories of Human Learning. Academic Press, New York, 1964.

AN ADAPTIVE PRIVATE PILOT CERTIFICATION EXAM

Stanley R. Trollip
Richard I. Anderson
University of Illinois at Urbana-Champaign

Every year there are millions of tests administered throughout the United States for a variety of reasons. In aviation alone, the Federal Aviation Administration (FAA) gives over 230,000 paper and pencil examinations annually. While these tests are generally regarded as being necessary, their cost of administration is enormous. Therefore, techniques for reducing these costs are appealing.

Two such techniques are administering examinations via computer and adapting examinations to each examinee's ability. This paper addresses the latter, specifically the application to aviation certification of adaptive testing based on item response theory. Adaptive testing is defined, item response theory is introduced, adaptive testing based on item response theory is described, and the impact of applying adaptive testing to the certification of private pilots is discussed.

Adaptive Testing and its Potential Benefits

An adaptive test can be defined as one which is constructed dynamically on the basis of an examinee's responses. In other words, many of the items to be administered are not selected beforehand, but are selected as the test proceeds in such a way as to extract the most information from the test. Thus it can be said that the test adapts to each examinee.

A test can be adapted several different ways. One way is to explore in depth only those areas in which the examinee demonstrates weakness. Another, which can result in or can be a part of the first, is to terminate a test as soon as the cutoff score has been or can not be attained. A third way, which can be a part of both the first and second, is to select test items which will differentiate best the ability level of the examinee from other ability levels.

Potential benefits are many. Because adaptive tests are often shorter than their conventional counterparts, cost in examinee and examiner time is reduced. Because items are selected on the basis of each examinee's responses, the precision of each estimate of examinee status can increase. Because examinees will encounter fewer items that are either too difficult or too easy, examinee anxiety may decrease and examinee alertness may increase. Because each test is likely to be short and different for each examinee, test security is strengthened.

Item Response Theory

An alternative to classical test theory (the theory which underlies the vast majority of current cognitive ability testing) is called item response theory (Lord, 1980). The major difference between these two theories is that item and test characteristics (for example, item difficulty, test reliability) are invariant in item response theory only. In classical test theory, the values of item and test statistics are dependent upon the group of examinees who responded to the test; that is, statistic values vary with the ability levels and the homogeneity of ability levels of the examinees upon whom the statistics are based. In item response theory, item and test statistics are not dependent upon the particular group of examinees who responded to the test; statistic values are constant for all groups of examinees. Therefore, item response theory allows one to describe very precisely the characteristics of an item or test before it is administered.

The basis of item response theory can be said to lie in the relationship between examinee ability and the probability of responding to a test item correctly.[1] For most items, the probability of a correct response tends to increase with examinee ability. However, this relationship is rarely linear. A graph of a common relationship appears in Figure 1.

As can be seen, the probability of a correct response changes most dramatically between ability levels x and y. Examinees between ability levels w and x are almost equally likely to respond correctly, and those chances are very slim. Similarly, examinees between ability levels y and z are almost equally likely to respond correctly, and those chances are extremely good. Clearly, this item would be of little value if the purpose of the test was to differentiate among examinees of ability levels w through x or y through z. However, if examinees between ability levels x and y were given many items of this type, the spread of the resultant scores would be very large; therefore, this item would be very valuable for differentiating among examinees between ability levels x and y.

Although the relationship between examinee ability and the probability of a correct response varies from item to item, most of the curves which characterize this relationship (commonly called item characteristic curves) possess somewhat of an S-shape. While retaining an S-shape, item characteristic curves often differ in three ways. Each is described in turn.

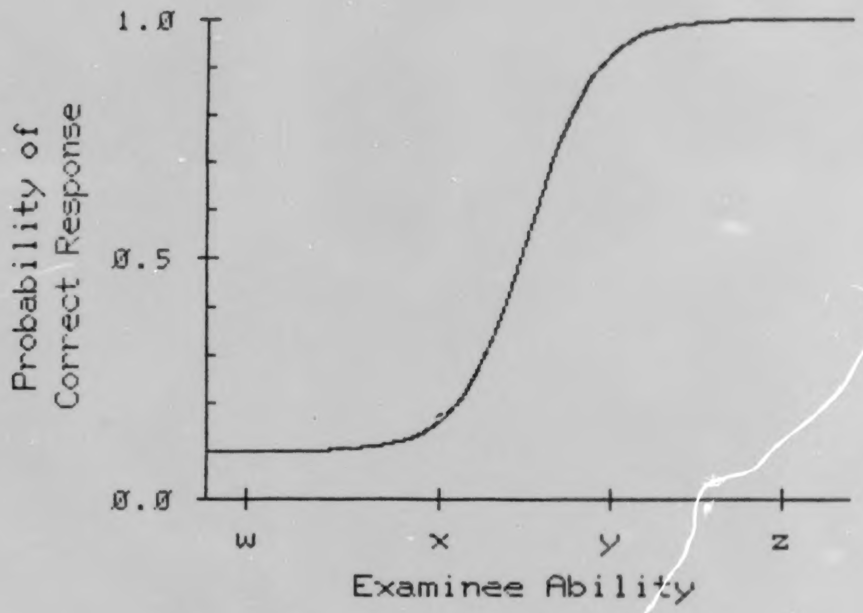


Figure 1
Common Relationship Between Examinee Ability and Probability of Correct Response

First, the point below which the curve never falls is item dependent. This point represents the probability that the "lowest" ability examinee will respond correctly. Not surprisingly, this point is commonly called the "guessing" parameter, and is most often near zero for constructed-response items and near the reciprocal of the number of alternatives for multiple-choice items. In Figure 1, the "guessing" parameter is 0.1.

Second, the steepness of the curve is item dependent. A steep curve, such as that appearing in Figure 1, reveals that the item will tend to discriminate very well among examinees within a very narrow range of ability. In other words, within a certain ability range, a steep curve reveals that the item will tend to discriminate very well between examinees that are close to one another in ability. As the steepness of the curve decreases, the power of discrimination extends over an increasing range of ability. However, the power of discrimination over this range is less than that over the narrow range of a steep curve. (Compare the two item characteristic curves in Figure 2 to verify this difference.) Not surprisingly, the parameter that represents the difference between the slopes of item characteristic curves is called the "discrimination" parameter.

Third, the position of the steepest slope of the curve

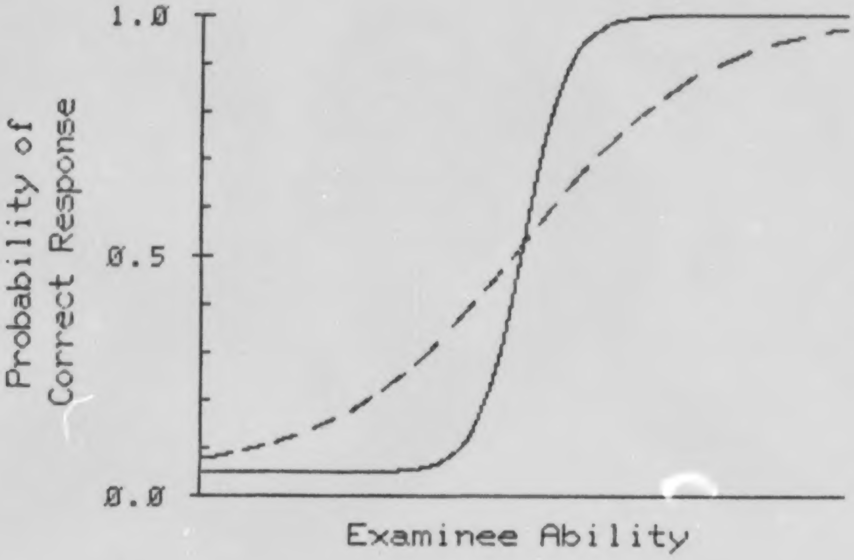


Figure 2
Two Item Characteristic Curves with
Different "Discrimination" Parameters

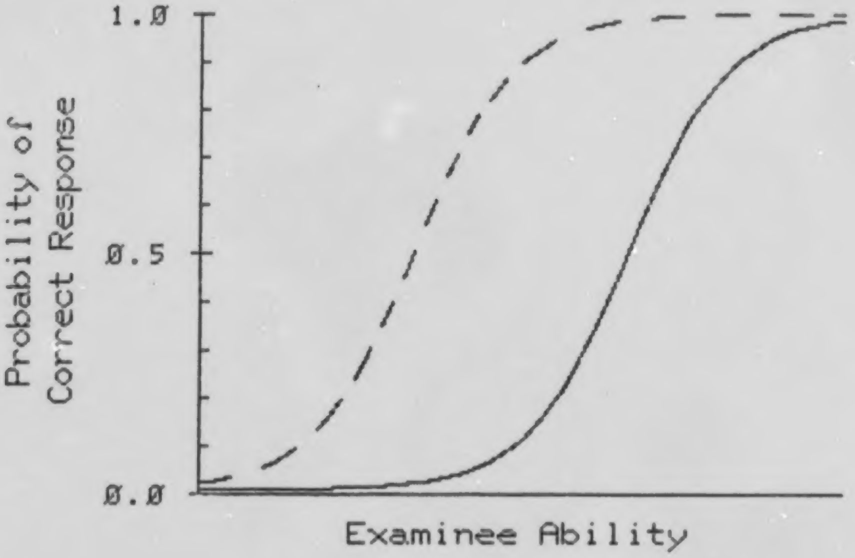


Figure 3
Two Item Characteristic Curves with
Different "Difficulty" Parameters

along the ability scale is item dependent (see Figure 3). This position is identified by the ability level at which an examinee's chance of responding correctly is midway between complete certainty and guessing. Although the difficulty of an item is actually examinee dependent, this position is commonly called the "difficulty" parameter. In Figure 3, the solid item characteristic curve represents an item that is more difficult for more examinees than the item represented by the broken curve.

The ranges of each parameter's values are parameter dependent. Of course, guessing parameters range from near 0 to near 1; according to Warm (1978), guessing parameters for multiple-choice items are typically lower than the reciprocal of the number of alternatives by 0.05. Discrimination parameters range from near zero to $+\infty$, such that the maximum steepness of the curve's slope increases with the parameter value; according to Warm (1978), values between 0.5 and 2.5 are most common. Difficulty parameters can range the entire ability scale.

Figure 4 presents the item characteristic curves of several items from the FAA's Private Pilot pool. These graphs demonstrate rather vividly the effect of different discrimination and difficulty parameter values on an item characteristic curve's shape. The guessing parameter of each of these examples is 0.2; truncation of the ability scale at the lower (left) end prevents some curves' appearance near 0.2. Truncation of the ability scale also prevents some curves' appearance near 1.0.

Adapting a Test to Examinee Ability Using Parameters as Defined by Item Response Theory

Once item parameters have been determined[2], items can be selected for administration to each examinee such that each selected item will contribute most to the estimation of examinee ability. This is possible because item selection is delayed, except initially, until after the examinee responds. After a correct response, the estimate of examinee ability is increased and a more difficult item is administered; after an incorrect response, the estimate of examinee ability is decreased and a less difficult item is administered. In general, changes in the ability estimates will eventually become progressively smaller. Once the change is negligible, the test can be halted. A schematic of a potential series of estimate changes appears in Figure 5.

Item selection is usually not based solely upon item difficulty. As discussed earlier, the steepness of an item characteristic curve reveals how much an item will tend to contribute to the discrimination among examinees within an ability range. After an estimate of ability has

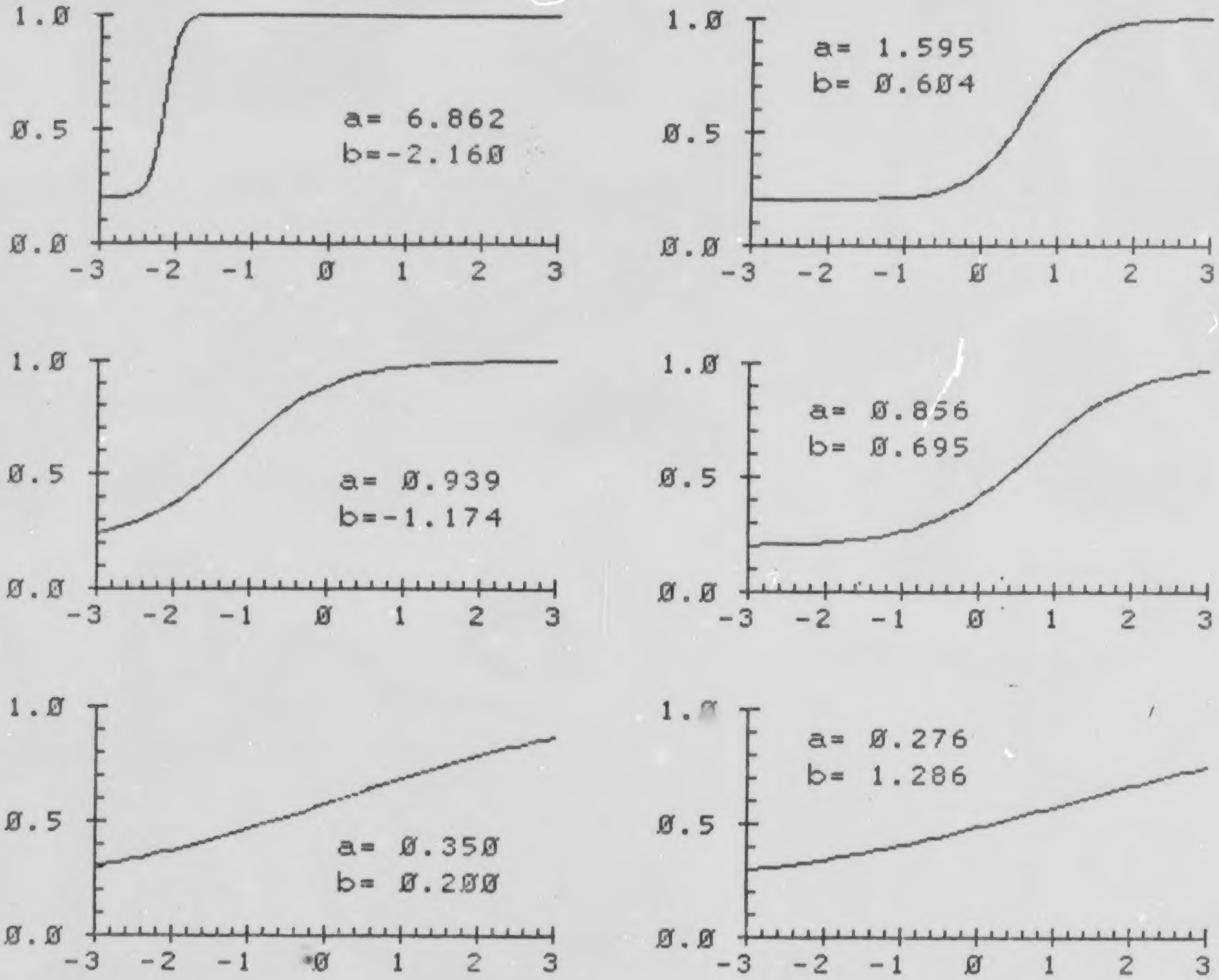


Figure 4
Characteristic Curves of Existing Items
(a = discrimination; b = difficulty)

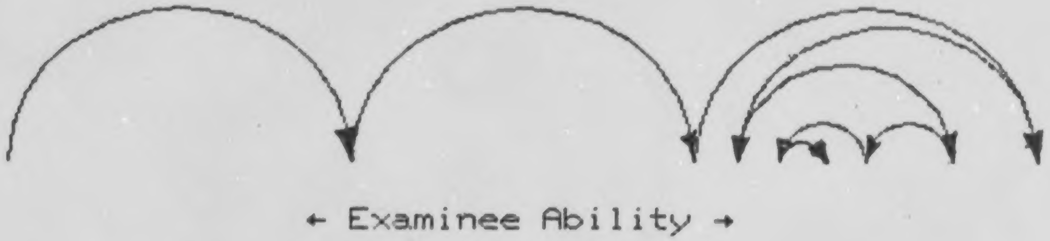


Figure 5
Potential Series of Changes in
the Estimate of an Examinee's Ability

been calculated, the item that will discriminate best between examinees at the estimated ability level and neighboring ability levels can be selected. Typically, this selection is such that the selected item's difficulty is greater than the previous item's difficulty following a correct response and lower following an incorrect response.

The estimate of examinee ability that is calculated after each response is such that an examinee of that ability has the greatest chance of generating the previously generated responses. Examinees of any ability level possess some chance of generating a particular series of correct and/or incorrect responses to a particular series of items. However, there is only one ability level at which the chance is greatest. Each estimate of examinee ability is the best approximation possible of this single ability level given the examinee's responses. In general, as more items are answered, the chance that an examinee at the revised estimate of ability has generated the generated responses becomes increasingly greater than the chances of examinees at other ability levels.

Adapting Private Pilot Certification

As part of a contract with the Federal Aviation Administration and the Office of Personnel Management, we are exploring the applicability to the FAA Private Pilot Written Examination of adaptive testing utilizing item parameters as defined by item response theory. As currently designed, the operation of such an adaptive exam will proceed as described in the previous section with only one difference. Since the primary intent of the private pilot exam is to differentiate those who should be certified from those who should not, termination of the exam can occur as soon as the estimate of examinee ability is precise enough to reveal that it is clearly below or clearly at or above the cutoff ability level. Thus, the exam does not always need to continue until a highly precise estimate of ability is calculated.

All of the potential benefits of adaptive testing that were discussed in general earlier in this paper are expected to accrue with an adaptive private pilot exam. In fact, since we are implementing the exam on a computer-based education system called PLATO (Lyman, 1980), an additional benefit is possible. Time saved because of a shortened exam can be used by an examinee to study computer-based aviation lessons or to respond to practice exams during which correct responses are revealed.

Despite expected benefits, the implementation of the adaptive test may cause a few problems. Some of these are discussed below.

One problem lies in the nature of feedback that is available from an adaptive certification exam. Currently each examinee is informed of the FAA categories in which appear the items he or she answered incorrectly. Though the quality of this feedback can be challenged because each category is usually represented on the test by a single item, this feedback is politically desirable and possesses legitimate benefit. With an adaptive test, however, feedback cannot be as extensive because many categories are likely to be excluded.

A particularly intriguing question surrounds the psychological and legal ramifications of failing an adaptive certification exam. Examinees may not accept failure when it is based on a shortened test. Therefore, it may be necessary to terminate a test early only when an examinee is assured of passing.

Lastly, because administration of the adaptive certification exam requires a computer, examinees must be willing to take the certification exam via computer. In an attempt to determine if private pilot examinees dislike a computer-administered exam, we have implemented a mimic of the current private pilot exam on PLATO and are using it to examine student pilots officially. Any student pilot wanting to take the FAA certification test locally is given the option to take our computerized version. Data to date indicate that the computerized version causes no logistical problems in terms of use and is well received, particularly because of the immediate feedback. However, it also appears that if an examinee is dubious about passing, the paper and pencil test is preferred, possibly because the immediate notification of results could cause embarrassment. Such discomfort could contaminate examinee performance on a computerized version.

Implementation of the adaptive test may not only cause a few problems, it may also be a problem. A major obstacle lies in the potential multidimensionality of the private pilot item pool. A single adaptive test using item parameters defined by item response theory cannot be implemented unless the subject matter covered by the test is all of the same type, in the sense that one ability is called upon to answer all the items. Our inspection of the private pilot pool suggests that it is multidimensional; that is, different abilities may be tapped by different subsets of items. If multidimensionality were to be confirmed empirically, the item response model could be applied only to subtests, conceivably diluting potential benefits. However, confirmed multidimensionality could reveal that some examinees can pass the current exam while failing on one or more critical dimensions. Thus a battery of tests may be more desirable than a single test. Application of adaptive logic to each subtest would still be possible, though the small size of subtest item pools may hinder the adaptability of some tests.

One obstacle that we encountered would not be encountered should the FAA decide to fully implement an adaptive private pilot exam. This obstacle involved our calculation of item parameters. Ideally, item parameters are calculated after the careful administration of the items to several hundred examinees. Alternatively, item parameters can be calculated from the responses of past certification examinees. However, since we were unable to implement either option, we were forced to estimate item parameters from classical test statistics. Unfortunately, the accuracy of such estimates has not been completely established, although existing evidence looks good (Jensema, 1976).

Conclusion

The increasing use of computers for the administration of tests opens new possibilities for improving the information available from tests and for decreasing the cost of giving them. One such possibility is the implementation of adaptive testing. This paper has described adaptive testing based on item response theory and has discussed several issues that are being considered during the application of this technique to private pilot certification. Despite unresolved problems, adapting aviation certification exams appears to be generally desirable and will probably be recommended in the specific case of private pilot certification.

Footnotes

1. As discussed by Warm (1978), the relationship of interest more precisely involves the amount of confidence (represented by a probability value) examinees have in the correctness of the correct response; it does not involve the probability of responding to a test item correctly. However, reference to the latter facilitates communication. Therefore, only the latter is used in this paper.

2. For a discussion about how item parameters can be calculated from examinee response patterns, see Lord (1980). For a discussion about how item parameters can be estimated from classical item statistics, see Schmidt (1977).

References

Jensema, C. A simple technique for estimating latent trait mental test parameters. Educational and Psychological Measurement, 1976, 36, 705-715.

Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, N. J.: Lawrence Erlbaum, 1980.

Lyman, E. R. PLATO highlights. Urbana: Computer-based Education Research Laboratory, University of Illinois, March 1980.

Schmidt, F. L. The Urry method of approximating the item parameters of latent trait theory. Educational and Psychological Measurement, 1977, 37, 613-620.

Warm, T. A. A primer of item response theory (Tech. Rept. No. 941078). Oklahoma City, OK: U. S. Coast Guard Institute, October 1978.

TOWARDS AN INTERNAL MODEL IN PILOT TRAINING

R. J. BRAUNE S.R. TROLLIP

AVIATION RESEARCH LABORATORY
UNIVERSITY OF ILLINOIS

INTRODUCTION

Research into the problems of pilot training has been dominated for a long time by the suggestions coming out of behavioristically oriented research. However, these suggestions have not always led to the expected success. Current findings in cognitive psychology and human information processing together with the changing role of the pilot seem to suggest that a new approach to flight training would be justified.

One of the areas where the behavioristic approach did not lead to the desired success was in instrument flight training. In accordance with the Stimulus-Response model of behaviorism the student pilot is supposed to develop a rigid scan pattern in order to extract the necessary information from the flight instruments. A long line of research showed however that experienced pilots do not follow a rigid scan pattern when flying under instrument conditions (e.g. Jones, R.E., Milton,J.L., Fitts, P.M., 1950; Milton, J.L., Jones,R.E., & Fitts, P.M.,1950; Gainer, C.A. & Obermayer, R.W., 1964; De Maio, J., Parkinson, S.R., & Crosby, J.V., 1978; Spady, A.A., 1978; Simmons, R.R., Lees, M.A., & Kimball, K.A., 1978; Harris, R.L., & Christhilf, D.M., 1980; Tole, J.R., Ephrath, A., & Young, L.R., 1980; Dick, A.O., 1980). The research mentioned has been reviewed in detail by Braune (1981). The empirical evidence seems to suggest that experienced instrument pilots adjust their scan pattern to the requirements of a given situation with the attitude information as the central source. Although individual differences exist in the sequence in which pilots scan the instruments in a given situation, it appears that pilots - in the same situation - look for the same categories of information. It is being proposed that scanning is controlled by cognitive factors, specifically, task demands. Through experience the pilot has developed an internal representation of flight phases which generates certain expectations about given situations. These expectancies guide the pilot's scan. When scanning the instruments the pilot develops a mental picture of the current situation. If the mental picture correlates with the expectancies no adjustments will be necessary. Otherwise control inputs will have to be made.

THEORETICAL RESEARCH

The idea of an internal representation with built-in expectancies has not been limited to aviation. Based partly on findings that came out of applied research it was attempted to develop theoretical models that would help to explain and to predict the monitoring behavior of human operators.

Work by Hickey & Blair (1958), Senders (1964, 1965), Carbonell (1964), Smallwood (1967), and Sheridan (1970) led to Moray's (1976) discussion of an internal model in human operators. Moray suggests that the data coming out of structural and functional research into operator performance may be combined under the concept of an internal model. The growing acceptance of the idea of an internal model seems to be expressed in a number of papers that were presented at the NATO Symposium on "Human Detection and Diagnosis of System Failures", Fall 1980 (Bainbridge, 1980; Curray, 1980; Rasmussen, 1980; Wickens & Kessel, 1980; Moray, 1980; Epraht & Young, 1980).

The general ideas involved in the internal model of human observers may be summarized as follows:

- experienced operators have an internal model of the overall system. The internal representation of the system's processes is not limited to visual cues, but includes a wide range of sensory stimuli. Any one of the stimuli that is integrated into a high fidelity internal model can guide the operator's control behavior.
- an important part of the internal model is a thorough knowledge and understanding of the system's dynamics. This allows the operator to have certain expectancies. These expectancies allow for a reduction in workload.
- the knowledge of the system's dynamics is the primary guide for the information sampling behavior of the operator concerning the state of the system.
- the sampled information is being used to construct a mental image of the current state of the system.
- the mental image is being compared to the internal model, or the expected state of the system in order to make a decision about control actions.
- the control actions are directed towards the future state of the system which lies within the goals determined by the internal model.
- the control actions consist of response schemata which enable the operator to adapt to the situation.

The use of pilot scanning models and the correlated idea of an internal model in the training and evaluation of pilots could prove to be a significant contribution to flight training, especially if it is being considered that no real understanding concerning efficient training methods exists.

CURRENT TRAINING METHODS

The idea of stimulus-response pairing and the measurement of observable behavior was readily accepted in the training environment. It was probably Gagné's work (1970) that was most influential concerning the analysis and classification of observable behavior. Gagné (Gagné, 1970; Gagné & Briggs, 1974) shows how -- under the concept of "task analysis" -- learning objectives can be developed which supposedly lead to the mastery of the skills involved in a given task. The current instructional systems development technique (ISD) of the United States Air Force is based on this approach. Miller (1974) and Klein (1977) criticize this approach with the argument that it does not handle certain training issues efficiently, for example complex non-procedural tasks.

Another major problem with the current approach to training is that research into the problems of how to train a pilot has been neglected. To the authors' knowledge only very few attempts have been made to build an information base concerning the cognitive and information processing aspects of flight training (Goldstein & Goldstein, 1972; Prather, 1973; Trollip, 1976). It appears as if no systematic framework exists which could guide the research into the training problems of future pilots.

PROPOSED FRAMEWORK

Considering the suggestions centering around the internal model of human operators one of the major questions of future research dealing with training techniques should be how the internal model can be developed efficiently and correctly. We suggest that an answer to this question may be found if the representation of knowledge is considered.

In recent years a growing number of scientists in human information processing, cognitive psychology, and artificial intelligence have been working on the problem of the representation of meaning and the structural and processing aspects of knowledge (e.g.: Bobrow & Norman, 1975; Minsky, 1975; Rumelhart, 1975; Schank, 1978; Anderson, 1980). Although frequently differing in the vocabulary being used the general idea is the same, namely that stored knowledge is being organized in the form of interactive knowledge structures called "schemata".

Rumelhart & Ortony (1977) define schemata as "data structures for representing the generic concepts stored in memory. They exist for generalized concepts underlying objects, situations, events, sequences of events, actions, and sequences of actions...". Nelson (1977) defines a schema as "an event sequence that describes the interaction of a number of concepts organized around a goal. Knowledge of schemata for recurrent events enable to predict what, when, and why things happen".

Schemata have variables that may become associated with different aspects of the environment on different occasions. For example the schema for LANDING has at least 3 different variables: a landing site, a particular type of aircraft, and certain environmental conditions. On different occasions these variables will take on different values.

Schemata contain as part of their specifications information about the type of values that may be connected to the variables. This serves at least two important functions: a) they tell what range of values might realistically be bound to each variable; b) when there exists insufficient information they can allow for good guesses and expectations for at least some of the variables. The constraints are seldom absolute but represent statistical distributions of possible values.

The general structure of schemata appears to be given in terms of relationships among other schemata. Schemata can be embedded in one another. This feature allows one to understand objects or situations in terms of their major parts without necessary reference to the internal structure of the constituents themselves. Embedding also seems to suggest a representational economy of variables. A subschema that appears in every LANDING is the DESCENT. However, this subschema is not limited to the LANDING schema but can also be utilized during any other phase of the overall schema of FLIGHT.

Rumelhart & Ortony (1977) suggest that there exists a close relationship between plans and actions. Actions can perhaps be viewed as schemata instantiated with motor values. That the application of schemata to motor behavior seems possible has been shown by Schmidt's (1975) extension of Adams' (1971) closed loop theory of motor learning, the schema theory of discrete motor skill learning.

Schemata may be considered to be the key to the comprehension process. Comprehension may be considered to consist of selecting schemata that will correlate with the perceived information, and then based on a decision rule verify that those schemata account for it. A schema correlates with the perceived information whenever that information can be interpreted as an instance of the concept that the schema represents. Looking at it from the point of view of an internal model this would mean that the perceived information is used to construct a mental image. The mental image is then being used to establish a correlation with the internal model of the given situation.

Probably the most important aspect of the schema approach to comprehension is the role of prediction. Assuming that the process of comprehension can be regarded as being similar to the process a scientist goes through in testing a theory, it is reasonable to agree that not every possible experiment must have been performed to be able to predict with some confidence the outcome of many proposed experiments. Schemata permit one to predict aspects of the input which have not yet been and perhaps never will be observed. For instance, once a pilot has rolled his aircraft into a turn he can expect his aircraft to loose altitude because of the overall reduction of lift. The schema TURN consists of certain aspects - certain variables - which allow the pilot to make predictions and act accordingly.

The particular schema which will be activated at the time of perception depends not only on the input itself but also on the context in which it occurs. Different contexts may give rise to different patterns of schemata available for comprehension even though the input might be the same (Bransford and Johnson, 1972). It is assumed that no random search procedure could possibly lead to an efficient activation of schemata. The search for possible schemata has to be guided and it must be sensitive to the context at hand. The same input is interpreted differently by observers depending on the conditions under which they observe it, what they have just observed, and what they expect to observe. What seems to be necessary is a process which allows for the convergence of information so that information derived directly from the input can be combined with expectations to activate the correct schemata. Rumelhart (1977) proposes that such a convergence can be achieved by combination of bottom-up and top-down processing.

The discussion so far has taken schemata as given cognitive structures that exist from the very beginning. No mechanism has been proposed yet whereby new schemata can grow and old ones evolve. The nature of schemata and their interrelation with concepts seems to suggest a number of mechanisms whereby new schemata can be invoked. Two of such mechanisms are schema specialization and schema generalization.

Schema specialization occurs when one or more variables in a given schema are held constant, for example the flight student learns to land one particular aircraft on one particular runway under the same environmental conditions. The criteria for schema specialization are probably frequency and utility. If a schema is used frequently with the same values assigned to some or all of its variables then the generation of a specialized schema will occur. Specific schemata allow for a fast detailed interpretation of a small range of inputs. Comprehension seems to be facilitated. However, if a person is confronted with too many specialized schemata the difference between them may not be sufficient to enable the selection of the correct ones. A discrimination problem may occur and the processing time saved in comprehension will be taken up in selection.

Generalized schemata are the exact opposite of specialized schemata. Some fixed portion of a specialized schema is replaced with a variable to construct a more abstract schema. Carefully selected examples of a given schema will help to demonstrate the nature and limits of values that its variables can take. The aqcuisition of generalized schemata is especially desireable when dealing with situations in which no one schema fits exactly - the type of situations that can be found in a complex, dynamic, high performance environment like aviation.

CONCLUSION

The purpose of training, especially in the area of aviation should be to provide a skill and knowledge structure that will prove useful to a pilot in processing new information and dealing with novel situations. However, it was mentioned that the current approach to flight training is derived from Gagné's recommendations which assume that: a) learning progresses from simple to complex; b) skill learning is additive (operant conditioning); c) a given skill exists in a clearly defined problem space (Stimulus-Response pairing). In the highly dynamic environment of aviation it appears to be doubtful that especially the last assumption is valid. Instead it must be assumed that the pilot is confronted with a relatively ill-defined problem space that requires constant adaptations. adaptations. Considering the findings of pilot scan research and the theoretical issues centering around the notion of the internal model, it appears to be possible that the suggestions from schema theory may lead to a more efficient training of pilots. However, due to the nature of the recommendations given by schema theory, which were mostly derived in strictly controlled laboratory settings it is clear that they will have to be empirically validated within a systematic research framework in the aviation environment.

The following questions may serve as guidelines for research into the development of efficient training methods towards an accurate internal model.

- What concepts and rules should be included in a given schema?
  The context and the entry knowledge (conceptual and procedural) of the learner population will help in answering this question.
- What are the variables of a given schema and what is the range of the values of these variables?
  This question should be investigated in the different contexts in which a defined schema can occur. By studying an identified schema in these different contexts it should be possible to identify the variables and the range of values that they can take on.
- Will the acquisition of the internal model be facilitated if certain variables are held constant initially?
  The state of the art in simulator technology allows one to hold certain variables constant. When the flight student begins his training it is now possible to ignore certain conditions initially and have him concentrate on the aspects which suggest an easier integration into his existing conceptual and procedural knowledge structure. By reducing the learner's overall workload during the training process more attention can be given to the critical skills to be learned. Hence the intended automatization of the given skill should occur sooner (Shiffrin and Schneider, 1977).
- Can the acquisition of an accurate internal model be facilitated if some of the functional relationships are being trained context free before training in the specific context occurs?
  Current research into this area seems to suggest that this may be possible (Rouse, 1981).
- What considerations should be given concerning the examples which are supposed to communicate a given schema to the learner?
  a) Will the comprehension of the system dynamics be facilitated if the underlying knowledge structure has been taught explicitly, or does it suffice to present examples only?
  b) Will a divergent set of examples be better for the transfer to novel situations than a set of highly similar examples?
  c) Does the order of example presentations influence the acquisition of an accurate internal model?
  d) How many examples need to be presented in order to build an accurate internal model that facilitates the transfer to novel situations?

All these questions only represent an outline of the problems that need to be investigated. However, the lack of data concerning efficient training methods for the development of an accurate internal model suggestes that systematic research into these questions should be initiated. As the role of both military and commercial pilots changes, the need for this research becomes more pressing.

REFERENCES

Adams, J. A. A Closed-Loop Theory of Motor Learning. Journal of Motor Behavior, 1971, 3(2), 111-149.

Anderson, J. R. Cognitive Psychology and Its Implications. San Francisco: W. H. Freeman and Company, 1980.

Bainbridge, L. Mathematical Equations Or Processing Routines? Pre-Proceedings NATO Symposium on Human Detection and Diagnosis of System Failures, Roskilde, Denmark, August, 1980.

Bobrow, D. G., and Norman, D. A. Some principles of memory schemata. In D. G. Bobrow and A. M. Collins (Eds.), Representations and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Bransford, J. D., and Johnson, M. K. Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Braune, R. J. Towards an Internal Model in Pilot Training. Unpublished Master's thesis. Savoy, Illinois: University of Illinois at Urbana-Champaign, Aviation Research Laboratory, Technical Report ARL-81-1, 1981.

Briggs, L. Instructional Design. Englewood Cliffs, New Jersey: Educational Technology Publications, 1977.

Carbonell, J. R. A Queueing Model of Many-Instrument Visual Sampling. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 4, 157-164.

Curry, R. E. A Model of Human Fault Detection for Complex Dynamic Processes. Pre-Proceedings NATO Symposium on Human Detection and Diagnosis of System Failures, Roskilde, Denmark, August, 1980.

DeMaio, J., Parkinson, S. R., and Crosby, J. V. A Reaction Time Analysis of Instrument Scanning. Human Factors, 1978, 20(4), 467-471.

Dick, A. O. Instrument Scanning and Controlling: Using Eye Movement Data To Understand Pilot Behavior and Strategies. NASA Contract Report 3306, 1980.

Ephrath, A. R., and Young, L. R. Monitoring vs. Man-In-The-Loop of Aircraft Control Failures. Pre-Proceedings NATO Symposium on Human Detection and Diagnosis of System Failures, Roskilde, Denmark, August, 1980.

Gagné, R. The conditions of learning (2nd ed.). New York: Holt, Rinehart and Winston, 1970.

Gagné, R., and Briggs, L. Principles of Instructional Design. New York: Holt, Rinehart and Winston, 1974.

Gainer, C. A., and Obermayer, R. W. Pilot Eye Fixations While Flying Selected Maneuvers Using Two Instrument Panels. Human Factors, 1964, 485-501.

Goldstein, I., and Goldstein, G. Flying in your head. Cambridge, Massachusetts: MIT, Artificial Intelligence Laboratory, LOGO Working Paper No. 3, 1972.

Harris, R. L., and Christhilf, D. M. What Do Pilots See In Displays? Paper presented at the Annual Meeting of the Human Factors Society, Los Angeles, California, 1980.

Hickey, A. E., and Blair, W. C. Man As A Monitor. Human Factors, 1958, 8-15.

Jones, R. E., Milton, J. L., and Fitts, P. M. Eye Fixations of Aircraft Pilots, IV. Frequency, Duration and Sequence of Fixations During Routine Instrument Flight. USAF Technical Report No. 5975, Wright Air Development Center, Wright-Patterson AFB, Ohio, 1950.

Kantowitz, B. H., and Hanson, R. H. Experimental Results Concerning The Detection of Operator Failures in Display Monitoring. Pre-Proceedings NATO Symposium on Human Detection and Diagnosis of System Failures, Roskilde, Denmark, August, 1980.

Klein, G. Phenomenological Approach To Training. AFHRL-TR-77-42. Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory, 1977.

Miller, R. B. A method for determining task strategies. AFHRL-TR-74-26, AD-783 847. Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory, 1974.

Milton, J. L., Jones, R. E., Fitts, P. M. Eye Fixations of Aircraft Pilots, V. Frequency, Duration, and Sequence of Fixations When Flying Selected Maneuvers During Instrument and Visual Flight Conditions. USAF Technical Report No. 6018, Wright Air Development Center, Wright-Patterson AFB, Ohio, 1950.

Minsky, M. A. A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Moray, N. Attention, Control, and Sampling Behavior. Proceedings NATO Symposium on Monitoring and Supervisor Control, Berchtesgaden, West Germany, 1976.

Moray, N. The Role of Attention in the Detection of Errors and the Diagnosis of Failures in Man-Machine Systems. Pre-Proceedings NATO Symposium on Human Detection and Diagnosis of System Failures, Roskilde, Denmark, August, 1980.

Nelson, K. Cognitive Development and the Acquisition of Concepts. In R. C. Anderson, R. J. Spiro, and W. E. Montague (Eds.), Schooling and the Acquisition of Knowledge. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Prather, D. Prompted mental practice as a flight simulator. Journal of Applied Psychology, 57(3), 353-355, 1973.

Rasmussen, J. Models of Mental Strategies in Process Plant Diagnosis. Pre-Proceedings NATO Symposium on Human Detection and Diagnosis of System Failures, Roskilde, Denmark, August, 1980.

Rouse, W. B. Experimental Studies And Mathematical Models Of Human Problem Solving Performance In Fault Diagnosis Tasks. In J. Rasmussen and W. B. Rouse (Eds.), Human Detection and Diagnosis of System Failures, New York: Plenum Press, 1981.

Rumelhart, D. E. Notes on a schema for stories. In D. G. Bobrow and A. M. Collins (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Rumelhart, D. E., and Ortony, A. The Representation of Knowledge in Memory. In R. C. Anderson, R. J. Spiro, and W. E. Montague (Eds.), Schooling and the Acquisition of Knowledge. Lawrence Erlbaum Associates. Hillsdale, New Jersey, 1977.

Schank, R. C., Lebowitz, M., Birnbaum, L. A. Integrated Partial Parsing. Research Report No. 143, Yale Artificial Intelligence Project, 1978.

Schmidt, R. A. A Schema Theory of Discrete Motor Skill Learning. Psychological Review, 82(4), 225-261, 1975.

Senders, J. W. The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. IEEE Transactions on Human Factors in Electronics, Vol. HFE-5, 2-5, 1964.

Senders, J. W. A Re-Analysis of the Pilot Eye-Movement Data. IEEE Transactions on Human Factors in Electronics, Vol. 7, No. 2, 1966.

Sheridan, T. B. On How Often the Supervisor Should Sample. IEEE Transactions on System Science and Cybernetics, Vol. SSC-6, No.2, 1970.

Shiffrin, R. M. and Schneider, W. Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory. Psychological Review, 84(2), 127-190, 1977.

Smallwood, R. D. Internal Models and the Human Instrument Monitor. IEEE Transactions on Human Factors in Electronics, Vol. HFE-8, No. 3, 1967.

Simmons, R. R., Lees, M. A., and Kimball, K. A. Visual Performance/Workload Of Helicopter Pilots During Instrument Flight. USAARL Report No. 78-6, Ft Rucker, Al: U.S. Army Aeromedical Research Laboratory, 1978.

Spady, A. A. Airline Pilot Scan Patterns During Simulated ILS Approaches. NASA Technical Paper 1250, 1978.

Stephens, A. T., Tole, J. R., Ephrath, A., and Young, L. R. Pilot Eye Scanning Behavior As An Index Of Mental Loading. Cambridge, Massachusetts: Massachusetts Institute of Technology, 1980.

Trollip, S. R. The Evaluation of a Complex Computer-Based Flight Procedures Trainer. Human Factors, 21(1), 47-54, 1979.

Wickens, C., and Kessel, C. Failure Detection In Dynamic Systems. Pre-Proceedings NATO Symposium on Human Detection and Diagnosis of System Failures, Roskilde, Denmark, August, 1980.

THE TOMORROW LEARNING MACHINE

W. W. Castor
Aviation Simulation Technology, Inc.

ABSTRACT

The purpose of this paper is to discuss the need for the evolution of general aviation flight simulators into learning machines and to propose a specific set of functions and designs for the tomorrow learning machine.

The term "flight simulator" is used generically to represent the gamut of training devices other than aircraft that are used in the flight training exercise. These include both generic and type simulators, flight training devices, and cockpit procedure trainers.

The term "learning machine" applies to devices that are interactive to students in the pursuit of training objectives. They are represented today by such tools as computer assisted instruction (CAI), "smart" video tape or disk and various mechanical tools that force a reiteration of learning, based upon the correctness of a student's response.

The concept of the tomorrow learning machine is the marriage of simulation techniques to computer controlled learning machines.

The discussion will be represented in four parts:

1. That there will be an increasing use of simulators in general aviation flight training.
2. That there are a number of inherent problems in flight training today.
3. That there is the potential for resolving these problems through the use of simulators that are adapted to become learning machines.
4. And that there is a proposed set of functions, uses and design criteria to be considered for the tomorrow learning machine.

THE GROWTH OF SIMULATION

General aviation training will be moving rapidly in the direction of simulation devices. The advanced experience of commercial and military aviation in the use of simulators has proven the benefits of increased proficiency for less cost.

And there is a major set of demands that exist today in general aviation that will increasingly drive towards training in simulation devices.

ECONOMICS - Flight training in simulators is less expensive.

ENERGY CONSERVATION - Simulators save on our energy reserves. You can literally fly around the world on the equivalent of one gallon of fuel.

EFFICIENCY - You can acomplish three to ten times as much in a given unit of training time. An aircraft is a very poor classroom by comparison.

EFFECTIVENESS - A greater range of training can be accomplished; particularly in handling emergency procedures. Simulators make safer pilots.

STANDARDS - Simulation provides an improved vehicle to standardize training...and simulators reduce the demand on air traffic control.

In short, simulation can be a superior way of handling training. And our technology today allows us to provide increasingly sophisticated elements of simulation at a price affordable to general aviation.

Projections of growth in simulator training in this decade range from 50% to 100% of all flight training being done in simulators by the year 1990. One large flight school is implementing plans to be doing 70% of their flight training in simulators within three years.

The tools of simulation in flight training will see extensive growth and usage. The numbers of general aviation simulators will increase one hundred fold by 1990.

With this dramatic proliferation of simulation devices and with the extraordinary wealth of technology we have today, how can we take further advantage of the opportunity to improve the effectiveness and efficiency of flight training

LIMITATIONS OF FLIGHT TRAINING

There have been traditional limitations to flight training since its inception. For the last 78 years flight training has been a one on one operation. One instructor and one student. It has been an event orientation.

As observed by Stanley Roscoe, in his book Aviation Psychology, "there is more literal truth than hyperbole in the frequent assertion that the flight instructor is the greatest single source of variability in the pilot training equation"....."the standardization of flight training has always proved an elusive objective".

We are not in fact able to standardize.

And because we cannot fully control or evaluate this variable event represented by the flight instructor, we effectively do not really know what works best in flight training. Nor are we quite sure about what is not working at all.

Education in flight training is not an event, it is a process. And to manage a process, we must specifically assign and calibrate the events. Because the events of flight lessons have great variability, we are woefully ill-equipped to implement and manage the process of education in flight training.

We have seen heroic and monumental efforts to improve flight training. By "improve" I mean elevate the success record as perceived by at least two terminal evaluations:

1. The accident rate per 100,000 hours; and ....
2. The check pilot. Another subjective and variable event.

Many of these improvement efforts have been represented by massive increased dosages of training and enormously expensive physical replicas of aircraft that are used as training machines. And they have had a measure of success.

We have made positive inroads. We have improved. But we really haven't done anything of substance to control and standardize the event! We've just added more of them.

Moreover, I would suggest that there may be more effective and less expensive ways. That we may not need to dump fertilizer over the entire farm in order to grow ten strawberry plants. If we could just target in on the strawberry plants, we could save a lot of fertilizer costs. Maybe we would even use a different kind of fertilizer.

But in order to rifle in on the specific elements of training; and in order to evaluate the training standards; and in order to eventually implement and manage a process of education... we need to standardize and empirically evaluate the event.

And with the assumption that simulation and flight training devices will be de rigueur in future flight training...we just may have a better opportunity to standardize and evaluate.

THE NEEDS FOR IMPROVEMENT

What are some of some of the obvious needs in flight training today that could potentially be addressed through the use of simulators?

There is a need for:

1. Standardization tools that can be interactive to a student's individual goals and progress.

There is a need for:

2. A means of augmenting an instructor's role to reduce the boredom of monitoring and to allow him tools to manage exceptions in the training process. To monitor the process of education as opposed to originating an event.

There is a need for:

3. A way of permanently recording flight training sessions to allow empirical comparisons of lesson effectiveness with a student or a group of students over time.

There is a need for:

4. Vehicles that force the implementation of consistent performance standards. For example: what is a successful VOR tracking lesson? If you ask 10 instructors, you will receive 10 different answers. We need to evaluate successful performance on specific parameters that are consistant among students.

There is a need for:

5. Increased instructor productivity. Why one on one? Why not one to 5... or 10?

FUNCTIONAL ELEMENTS OF TOMORROW'S LEARNING MACHINE

If we accept that flight training is moving rapidly in the direction of simulation tools, and that there are these needs for improvement in flight training as it exists today, then how can we endow simulators with the capacity to better resolve these needs for improvement? How can we interlace interactive training machines with simulators to serve as a platform for increased training proficiency? What might be proposed as the functional elements of tomorrow's learning machines that can contribute to the resolution of these needs? and that can allow the further generation of improved training techniques and tools.

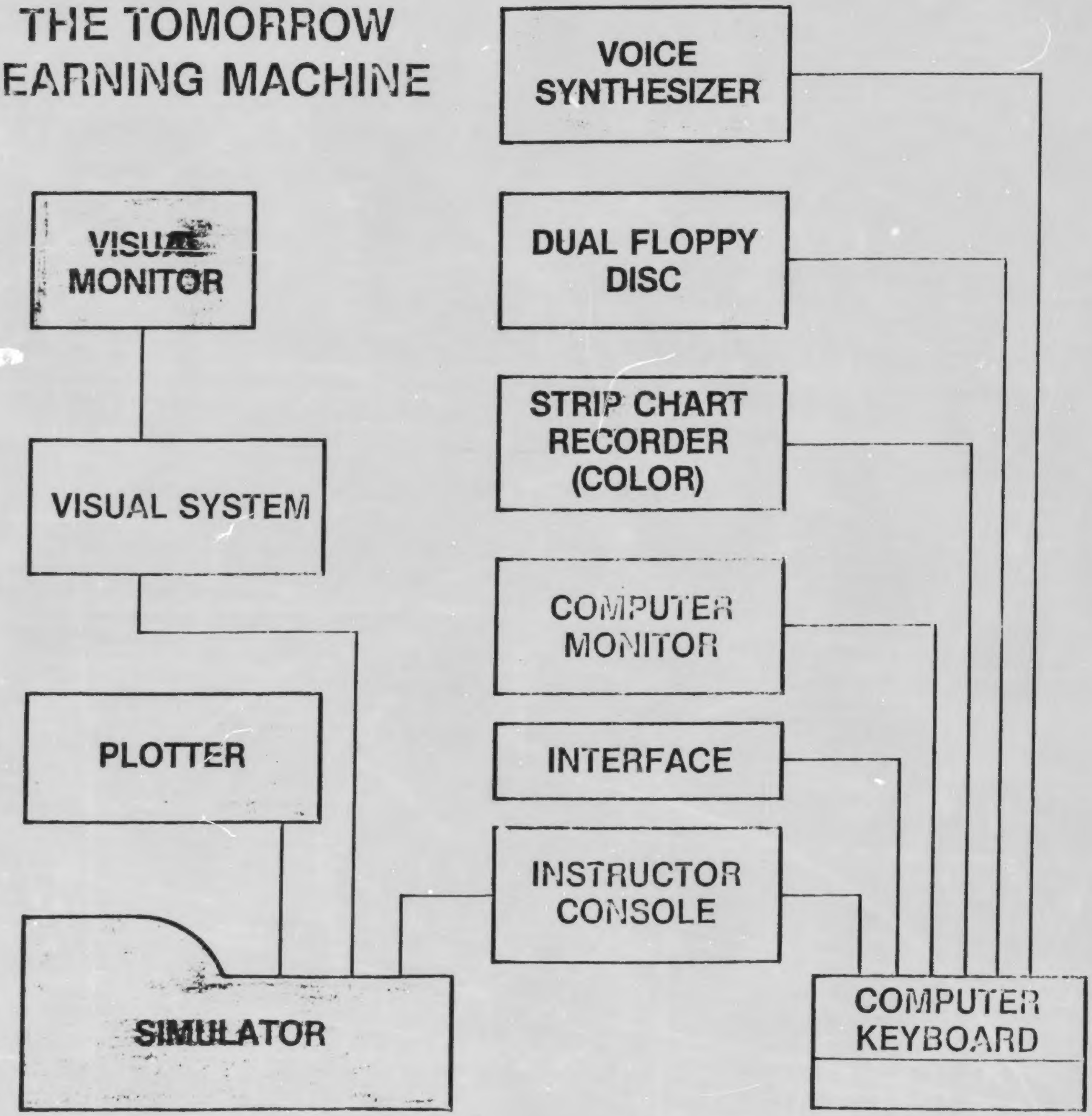
The following functional elements for tomorrow's learning machines are proposed for consideration:

1. For standardization: the implementation of simulator lessons at the interactive computer level. By that is meant that the training activities and parameters would be directed and monitored by computer programs. And that they would be designed to interact with the student by computer instruction to direct the lesson in the context of the student's performance. For example, if the student does not perform an ILS within prescribed parameters, he is directed by the computer to redo it. If the student's performance is extensively beyond parameters, an instructor is notified. The instructor would at that time bring his skills to bear to assist the situation. He is keyed to manage the exception. He is not required to sit and wait for an exception to occur. But the lesson would be standard. And the parameters of achievement would be standard. And the instructor productivity would be amplified.

2. Improved instructional tools are needed to increase instructor productivity and effectiveness. Rather than one instructor tending to one student, consider the use of remote centralized monitors that portray all lesson plan activity for several students. Monitors that could portray the track being flown as well as the attendant training parameters defined in each dimension of flight activity.

   There should also be the ability to recall and portray specific segments of the lesson by either monitor or by strip chart recorder to compare and reinforce the training experience as necessary.

THE TOMORROW LEARNING MACHINE


VOICE SYNTHESIZER
VISUAL MONITOR
DUAL FLOPPY DISC
STRIP CHART RECORDER (COLOR)
VISUAL SYSTEM
COMPUTER MONITOR
PLOTTER
INTERFACE
INSTRUCTOR CONSOLE
SIMULATOR
COMPUTER KEYBOARD

Emergency procedures should be reinforced in a standard fashion by pre-programmed or response generated emergencies to the instruments or to the engines. Because other students undergoing that same lesson would encounter a similar set of emergency conditions, it would be possible to evaluate a consistent level of training and training results.

Again, the instructor can amplify his productivity by monitoring many students at once. He would interject his skills only on an exception basis. he would be trained and equipped to manage the process of education.

3. Evaluation tools should be provided by permanently recording the lesson activity on computer media. It would be recorded in each dimension along with the specific parameters of required performance.

   This would allow, finally, an exact and precise record to evaluate the effectiveness of lessons, the longer term performance of students, the comparative effectiveness of various lesson plans on groups of students and an evaluation of the training device itself. It could allow a platform of knowledge upon which to base the evolution of behavioral training techniques and tools.

DESIGN REQUIREMENTS

In order to project a more defined reality to these functional elements, let us consider the general design and interface requirements for the components of TOMORROW'S LEARNING MACHINE.

The essence of what you see in the diagram (Figure 1) is a simulator with its component peripherals of visual system, plotter and instructor control console interfaced to a computer that uses a voice synthesizer, a dual floppy disc, a color strip chart recorder, a monitor, and a keyboard for its component peripherals.

The simulator, either generic or type specific, would require a databus and a capability to regularly emit precise digital information on attitude, position, altitude, speed, and condition. It further would require input ports to allow elements of external direction and control.

The interface would provide information to the computer from the simulator databus and allow the computer to control the visual presentation and the remote instructor console. The visual should be controlled in terms of the ability to dynamically change the ceiling, the visibility, the day or night toning and the runway light intensity. The instructor console should be automatically controlled for the implementation of various emergencies on a pre-programmed or response-oriented basis.

On the computer side, the voice synthesizer would be used to direct the lesson activity and it should have the capability of several different voices for its vocabulary. This would imitate the realism of receiving instructions or clearances from different clearance authorities. For example, one voice for the tower, another for departure control, and yet another for center control.

The dual disc would be for the purpose of holding lesson programs and for recording the flight lesson. The lesson plan and the parameters of performance would all be contained on the disk along with the ability to direct the lesson and to direct an iteration of a poorly performed maneuver.

The strip chart recorder should use color to allow distinct portrayal of the lesson track versus the actual flight track in each dimension.

The monitor (CRT) is there to provide lesson reinforcement. It is in effect an automated blackboard which could visually present a lesson objective by showing historic track or demonstration pictures.

The keyboard would of course be present for logging in and off and initializing programs....as well as providing instructor control and override.

The system could be further interfaced by modem to a central host computer or to a time shared system. These features could expand the learning machine to multiple units tied to interactive video briefing stations for a total integrated learning system center. A larger host system could also allow the evaluations of more complex maneuvers.

In addition to what is visible in this diagram, there is a requirement for an operational program that provides for simple input and alteration of lesson plans with voice control, interactive branching, pre-programmed emergencies, control of the visual, lesson performance parameters and recording.

And the final and most important requirement is an array of highly disciplined and tested lesson programs. And that will be a limitless goal because the learning, teaching evaluation is heuristic by nature. It is a closed system of specific requirements and exact behavioral results that will lead to continuing improvements.

Some of the specific capabilities of this learning system design would be:

1. To record altitude and position for a one hour flight.
2. To replay parts of the flight onto a strip chart with a horizontal situation that shows the allowable position criteria and a vertical cross-section showing the glide slope.
3. Instantaneous position on the computer monitor to orient the simulator position to facilities.
4. Computer generated instrument failures.
5. Instructions stored on floppy disc.
6. Programs that could be created by a user rather simply.
7. The ability of the computer to talk by using a stored vocabulary.
8. Computer generated instructions and corrections when a student is off course.
9. And computer generated check lists to reinforce procedural training.

And some of the functional uses of this specific design would be:

1. Computer interactive instruction where the computer provides instruction by voice and concurrently evaluates a student's performance on the simulator. The evaluation criteria would direct branching in the lesson for repeat maneuvers or for advanced progress.

2. Flight profiles where a typical trip begins with the copying of a clearance from the computer. Controllers' voices are changed and there are frequency hand-offs. Again, the computer would observe and record the performance. The computer could generate instrument failures at critical times or in response to a student's situation. The computer could control visibility and cloud ceiling to give enroute weather changes.

3. Instructor evaluation packages would allow the instructor to call for various evaluation plots or to request pre-programmed demonstration pictures.

4. Complete flight profiles would be permanent records. They could be used to evaluate student/group performance against a consistent standard...and they could be used to establish improved training profiles.

That then is a generalized picture of a proposed design for tomorrow's learning machine. The inherent capabilities of such a design could be instrumental in meeting the needs for improvement in flight training.

It would allow us a vehicle to standardize, control and measure what has been for the last 78 years the one great variable in flight training: the individual lessons by the individual instructor.

SUMMARY

There will be an increased dependency on, and requirement for, simulators in general aviation. The proliferation of simulators will grow exponentially. The future is moving towards us very rapidly.

There are pronounced needs for improvement in flight training today. But until we can get our hands around and standardize the event of a training lesson, we cannot control or improve the training process.

This trend towards significantly increased usage of simulation devices provides us with a unique opportunity to meet the needs for improvement in flight training today. And that opportunity can best be exploited by the marriage of simulators to interactive learning techniques. That marriage would provide an increased ability to standardize, control and measure the events of flight lessons. It would provide improved instructional tools that increase instructor productivity and concurrently provide the ability for students to proceed at their own pace and according to their own goals.

We have discussed at some length the functional elements, the design characteristics, the capabilities and the uses of tomorrow's learning machine. The technology of today makes feasible these characteristics. We are on the edge of a new renaissance in flight training.

We are directing our time and efforts to develop such a tomorrow learning system. We believe that the benefits of improved instructor productivity, standardization of training and especially the ability to empirically evaluate the effectiveness of various innovations in flight training techniques and tools can significantly advance the art of training to the level of a science.

We would very much appreciate the input and observations of others as to how we can direct and amplify the value of what we are trying to do.

OPERATOR SKILL RETENTION IN AUTOMATED SYSTEMS

Dennis B. Beringer
Department of Industrial Engineering
University of Wisconsin at Madison

Many contemporary man-machine systems are becoming increasingly automated. In some instances this automation is virtually complete (robot-operated assembly lines) while in others it is only partial (commercial aircraft autopilots). Although very high levels of reliability have been achieved in the operation of many such systems, some experience periodic failures ranging from slight performance degradation to catastrophic malfunction. The behavior of the human operator during these periods of failure is of particular interest.

Numerous semi-automated systems retain the option of direct human control although the operator may spend most of an operational cycle as a monitor of the system rather than a manipulator of the system. This essentially removes the operator from the control loop and poses serious problems when a manual override of the system becomes necessary. Beyond the immediate difficulty of knowing what state the system is in when such action becomes necessary, the operator who is out of the control loop for any length of time may also experience a reduction in manual control skills over time. This may be most pronounced if system failures are very infrequent.

The program proposed herein will examine the effects of task automation on skill retention through the use of simulaton. A PDP-11 computer will be used to generate displays and monitor operator performance on simulated monitoring and controlling tasks. Automation failures will be systematically introduced and will vary in frequency between experimental conditions. Performance under these conditions will be compared with that obtained using an artificial manipulation task designed to maintain skilled performance in the absence of actual system failures. This approach should provide a means of predicting what level and type of interpolated activity may be substituted for actual system control activity to maintain operator skills at an acceptable level.

MEASURING PILOT AIR COMBAT MANEUVERING PERFORMANCE

Edward E. Eddowes
Operations Training Division
Air Force Human Resources Laboratory

The key requirement in developing measures of any human performance is to know and understand the performance to be measured. In the case of pilot air combat maneuvering performance this requirement has become a formidable problem from the outset. While previous attempts have been made to know and understand pilot performance in air combat for the purpose of developing suitable measurement procedures, to date none have solved the problem effectively to the satisfaction of behavioral scientists and experienced air combat pilots.

The Mystique of Air Combat Maneuvering Performance

It is not surprising that acquiring the needed knowledge and understanding of a pilot's air combat performance is not easy. The mystique which surrounds the air fighter's skilled performance reinforces a popular notion of the courageous, determined, relentlessly aggressive, sharp-eyed hunter of the skies, capable of making split-second decisions and lightning-fast maneuvers to defeat an aerial opponent. Such images as these have been fostered in the film performances of Errol Flynn, David Niven, John Wayne, Robert Ryan, Edmund O'Brien, Robert Stack, Robert Mitchum, and Robert Wagner, to mention only a few.

The descriptions of fighter pilots presented by Tom Wolfe (1979) in The Right Stuff suggest such additional characteristics as ambitious, competitive, jealous, exploitive, and above all eager to accept dangerous challenges in order to demonstrate courage. A pilot with the right stuff will let it all hang out, against all odds, and will compete at the drop of a hat for the opportunity. Furthermore, righteous pilots don't talk about their right stuff. Thus the problem of knowing and understanding pilot air combat maneuvering performance to develop the means for measuring it remains unsolved.

Until recently further research and development which focused on pilot air combat maneuvering performance was discouraged. It was assumed to be too amorphous, too multi-faceted, too complex, too subjective, and too rare to attack systematically with any reasonable chance for success.

What Can We Learn from the History of Air Combat?

There are historical data available to aid in the search for knowledge and understanding of pilot air combat maneuvering performance. If one keeps a sharp, fighter pilot-like, eye out for them, more than 100 paperback histories, biographies, autobiographies and memoirs published/reprinted during the last twenty years may be found in drug stores, airline terminals, news stands, and book stores. From these histories of Rickenbacker (1969), Bishop (1967), Fonck (1967), and von Richthofen (1969) from World War I, and of Galland (1965), Hartman (Toliver and Constable, 1971), Tuck (Forrester, 1978), Townsend (1969), Sakai (1965), Klostermann (1979); Johnson (1964), and Scott (1965), from World War II, one can piece together an emerging picture of what a fighter pilot is, how he goes about his combat flying job and the behavioral processes he exercises in accomplishing an air combat mission.

The historical data at first underline the popular image of the expert air fighter as having superior vision (Sakai, 1965), as being an excellent pilot (Bishop, 1967) and deadly shot (Fonk, 1967; Heiferman, 1971; and von Richthofen, 1969), as exercising coolness under pressure, and above all as an aggressive attacker whose goal is to, "...blow you apart before you get close enough to hit me." (Scott, 1965). This last quote suggests a pilot characteristic that may provide a clue to help penetrate the fighter pilot mystique and stimulate a new interpretation of what the great aces have told us. Getting the enemy before he gets you suggests a strategy of taking no unnecessary risks, of fighting on your rather than his terms, of seeking and exploiting fighting advantages and, if possible, of avoiding contacts where your adversary may exploit an advantage against you. This approach has been endorsed by Fonck (1967), Rickenbacker (1967), Chennault (Heiferman, 1971), and Hartman (Toliver and Constable, 1971).

Such fighter leaders as Bishop (1967), and Chennault (Heiferman, 1971), indicate that the successful air combat pilot will learn everything he can about the capabilities and limitations of his aircraft and be practiced in extracting its maximum performance in a variety of air combat tactics and maneuvers under a wide array of environmental circumstances without unintentionally exceeding its performance limitations. For the same reason the expert air fighter will know the strengths and weaknesses of his adversary's aircraft, its performance capabilities and limits, and both the probable and possible combat tactics and maneuvers his enemy may use to counter his attack (Sims, 1962). In addition Fonck (1967) and Rickenbacker (1969) emphasize caution, concentration, and patience as key characteristics of the successful air combat pilot.

It can be seen that the expert pilot will plan and manage his combat engagement to put his opponent in the weakest defensive position by detecting his enemy before he is detected, for example, by hiding in the sun or behind clouds, so as to initiate an engagement from a favorable position. The expert will include in his plan a number of alternative tactics/maneuvers with which he can counter the enemy's defensive maneuvers and preserve his initial advantage as long as possible. Thus he will know ahead of time what to do, how to adapt what has to be done to the momentary circumstances of the engagement, what information he needs to guide him in optimizing the effectiveness of the attack, and where and when he must get this information, and what to do with it (Eddowes, 1974; Eddowes and Waag, 1980). This last bit of knowledge and planning foresight will prevent avoidable tactical errors during a fight, generate flexible criteria for moment to moment assessment of how it goes, and provide alternative courses of action for selection in case a pilot decides to knock-off his attack if it looks like an engagement is not going well enough to warrant pursuing it further.

Components of Air Combat Maneuvering Skill

There is one other important characteristic of the expert air fighter that may have little to do with aggressiveness, keen eyesight, or superior marksmanship. This attribute of air combat maneuvering performance involves the pilot's continuing application of his mission plan in sequencing and pacing his cockpit tasks in anticipation of what is coming up next. The result is his habit of getting everything done and out of the way as soon as possible in order to leave him relatively free to concentrate his attention on defeating his adversary once the engagement begins. All this reduces task overload and the performance degradation it typically produces.

The characteristic readiness-for-anything of an expert air fighter develops naturally as his flying skills increase in breadth, depth, and refinement during all the flying training he has undergone up to the moment he decides to point the nose of his aircraft toward his enemy. The pilot's habitual anticipation, that is, thinking ahead of the aircraft, has been trained-in since the beginning (Eddowes and Waag, 1980). Early in training a pilot learns that to control an aircraft properly, he must begin his control actions soon enough to satisfy the spatial and temporal requirements of the maneuver he is executing (whether it is leading with control stick pressure to level out of a dive without overshooting his assigned altitude, or starting a high yo-yo in time to avoid overshooting his adversary in air combat).

Acquiring satisfactory anticipation skill is complicated by the fact that errors often are not apparent to the pilot when they are made because of the lag in an aircraft's response to a control input. Thus a pilot can only review his error retrospectively to determine what he did wrong and discover how to correct/avoid the error in the future. This state of affairs does not provide favorable circumstances for feedback that leads to effective, rapid learning. Given that the pilot with average anticipation skill may need lots of luck to survive, much less succeed in air combat, it is easily seen that improved performance measurement could lead to more effective training and increase the rate of skill acquisition in air combat maneuvering.

Summarizing the analysis presented above, the air fighter has to translate his mission requirements into a planned sequence of actions he must take. Then he can use this information to evaluate his progress during the mission on the basis of how closely he satisfies the requirements of his plan. For example, by knowing the weapons loaded and how he will employ them, the pilot can prepare them for firing long before he is likely to contact the enemy. In addition, as a consequence of his planning for all foreseeable events, the air fighter can position himself in the best tactical position available to provide himself the most favorable escape routes or for modifying the mission plan if things turn out differently than expected. He will be able to evaluate the progress of an engagement on the basis of the criteria established in his mission plan. Further, he will be able to discriminate the need for deciding on what to do next because the plan will quide his information seeking which in turn will guide his decision making.

Applying Discrimination and Decision Skills in Air Combat

By simplifying his moment-to-moment performance requirements through anticipating the work he can accomplish before contact with the enemy, the air fighter can concentrate on the combat engagement and its critical cycle of discriminations and decisions that Hartman (Toliver and Constable, 1971) has identified succinctly, "See, decide, attack, break." The combat pilot sees or detects and identifies his adversary; decides to or not to attack, given the spatial relations involved, the functional state of his aircraft and the estimated capabilities of his adversary; attacks using his knowledge of his own aircraft and the enemy aircraft capabilities; and breaks contact following the attack using the most effective maneuver under the circumstances. Each engagement involves repetitions of the see, decide, attack, break discrimination-decision sequence. Like many of his predatory predecessors, Hartman found turning contests hazardous and avoided them. He preferred to use the see, decide, attack, break sequence to simplify his tasks in the midst of an aerial dogfight.

If this picture is accurate and valid, and the history of air combat indicates it is, the key components of an air combat mission are the quality of the mission plan and how a pilot uses it. The depth and breadth of the information the plan summarizes determines its quality, and the disciplined, refined discrimination and decision making of the pilot reflects the quality of his performance as he uses the plan to guide him in determining what to look for, where and when to look for it, and what to do with it when he finds it.

The air combat training problem, it seems, is organizing the environment of a novice air fighter so he can acquire the familiarity he needs in how to deal effectively with an air combat engagement rather than be overwhelmed by it. An important aspect of the training is feedback to the learner, such as may be provided by performance measurement information. The presentations which follow describe two independent approaches to developing air combat maneuvering performance measurement systems. They are based on different capabilities for recording descriptions of air combat engagements developed in different training environments by different military serivces. At the close of these presentations you will be able to *see* what has been accomplished, *decide* on the similarities and differences of these two measurement systems and applications, *attack* and assess the credibility and merit of the results, and determine if a *break* has been made in the previously mysterious area of measuring a pilot's air combat maneuvering performance.

References

Bishop, William A. Winged warfare. New York: Doubleday & Company, Inc., 1967. Pp. 175-181.

Klostermann, P. The big show. New York: Bantam Books, Inc., 1979.

Eddowes, E. E. A cognitive model of what is learned during flying training, AFHRL-TR-74-63, AD-A000 046. Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory, July 1974.

Eddowes, E.E., and Waag, W.L. The use of simulators for training in-flight and emergency procedures, AGARD-AG-248. Nevilly-Sur-Sien, France: Advisory Group for Aerospace Research and Development, June 1980. Pp. 4-5.

Fonck, R. Ace of Aces. Garden City, New York: Doubleday & Company, Inc., 1967. Pp. 124-125.

Forrester, L. Fly for your life. New York: Bantam Books, Inc., 1978.

Galland, A. The first and the last. New York: Balantine Books, Inc., 1965.

Heiferman, R. Flying tigers: Chennault in China. New York: Ballantine Books, Inc., 1971. Pp. 27-33.

Johnson, J.E. Full circle: The tactics of air fighting 1914-1964. New York: Ballantine Books, Inc., 1964.

Rickenbacker, E.V. Rickenbacker. New York: Fawcett Crest, 1969. Pp. 98-150.

Sakai, S. Samurai! New York; Ballantine Books, Inc., 1965. Pp. 24-25.

Scott, R.L. God is my copilot. New York: Ballantine Books, Inc., 1965. Pp. 121-130.

Sims, E.H. Greatest fighter missions. New York: Ballantine Books, Inc., 1962. Pp. 98-99.

Toliver, R., and Constable, T.J. The blonde knight of Germany. New York: Ballantine Books, Inc., 1971. Pp. 79-94.

Townsend, P. Duel of eagles. New York: Simon and Schuster, Inc., 1969.

von Richthofen, M.F. The red baron. New York: Doubleday & Company, Inc., 1969. P. 37.

Wolfe, T. The right stuff. New York: Bantam Books, Inc., 1979. Pp. 16-34.

PERFORMANCE MEASUREMENT AND THE NAVY'S TACTICAL AIRCREW TRAINING SYSTEM (TACTS) [1]

LT GERALD R. STOFFER, Ph.D
Human Factors Laboratory
Naval Training Equipment Center
Orlando, Florida 32813

ABSTRACT

The development and use of the Tactical Aircrew Training System (TACTS) as a means for training advanced air combat skills are described. Pilot performance measurement capabilities of the TACTS are reviewed in terms of their value for pilot selection, aircrew training, assessment, and simulator design. Several approaches and methods used to conceptualize and measure air combat maneuvering (ACM) performance are presented. Limitations in existing TACTS performance measurement capabilities are illustrated in terms of several psychometric, training, and TACTS operational user feasibility requirements for a system of performance measurement. It is concluded that while the existing TACTS represents a highly advanced aviation engineering technology that can provide extremely valuable training, that same technology has largely ignored the functional requirements for a system of human performance measurement. Improvements in the TACTS performance measurement capabilities will improve its training value even further.

INTRODUCTION

In the 1950 Korean conflict, American F-86 Sabre jet aircraft destroyed equally capable Soviet MIG-15 aircraft at a rate of 10 MIG's for every Sabre lost (Futrell, 1961). In view of these results, the USAF and USN were surprised at the relatively poor exchange ratios of approximately 2:1 advantage over the North Vietnamese Air Force during the first half of the Vietnam conflict (DeLeon, 1977; Editors, AFJI, 1974). In 1969, the Navy released the results of a study directed by Captain Frank Ault to explain the unexpected poor air combat maneuvering (ACM) performance of U.S. pilots in Southeast Asia. The "Ault Committee Report" (1969) identified deficiencies in air combat training as a primary factor. In particular, many pilots were reported to have fired their weapons (i.e., missiles) outside of tactical launch envelope boundaries. Failure to recognize and fire within an acceptable geometric cone surrounding a target aircraft greatly reduces the probability of kill. Because the Vietnam Air War was the first time that U.S. pilots used missiles as their primary air-launched weapon to destroy enemy aircraft in flight (Craven, 1980), inadequate training in missile envelope recognition was especially apparent.

In response to the Ault Report, the Navy developed a major requirement that a designated airspace (range) and ground support be designed and implemented to provide training in air combat skills. The specific range requirements were: (1) accurate real time weapon envelope recognition training during ACM; (2) accurate recording of training events for debriefing, and (3) safety and economy of ACM training (Air Test and Evaluation, VX-4, reported in Applied Physics Laboratory, 1975). The range, originally called the Air Combat Maneuvering Range (ACMR), is today known as the Tactical Aircrew Training System (TACTS). The general system, functional description, and operating guidelines for TACTS are described in detail elsewhere (User's Guide, 1978).

---

[1] This project has been supported by COMNAVAIRSYSCOM (340F) and NAVAIRDEVCEN (6021). The material presented here does not reflect official positions of either of these agencies or of the U.S. Navy.

PURPOSE

Two major training technologies exist to provide Navy air-to-air combat training. The first, TACTS, is the context for the present paper. Following a brief description of the TACTS operational and training capabilities, the remaining sections of the paper will focus on: (1) a description of ACM performance; (2) an analysis of the human performance measurement capabilities and limitations of TACTS; (3) an assessment of the value of TACTS performance measures for a variety of human factors applications; (4) an identification of different approaches used to conceptualize and measure ACM performance; and (5) development of a set of specific functional requirements for a system of performance measurement (PM).

The major purpose of this paper is to indicate the current limitations in Navy TACTS performance measurement relative to an ideal set of functional requirements for a performance measurement system (PMS). The major thesis developed herein is that while the TACTS original design included performance measures, there was no specific design for a system of measurement. It is felt that development of a PMS, which meets the requirements presented here, will greatly assist in more fully accomplishing the original operational training requirements which led to development of the TACTS facility.

The second major training technology, air combat maneuvering simulators (ACMSs), are somewhat more recent innovations. Two classes of simulators are used for Navy ACM training: (1) Weapon System Trainer, Device 2F112, located at NAS Miramar, CA and NAS Oceana, VA; and (2) Air Combat Maneuvering Simulator, Device 2E6, located at Oceana.[2] The ACMSs will not be discussed specifically below, except as an alternative approach to TACTS ACM performance measurement.

DESCRIPTION OF TACTS CAPABILITIES

Existing TACTS Operational Capabilities

TACTS is an advanced air combat training system developed to improve aircrew proficiency. In a designated airspace, controlled but realistic ACM missions are flown against U.S. "play to kill" aircraft adversaries that mimic the looks and tactics of Soviets MIGs. With special airborne instrumentation pods, all significant flight parameters, weapon events, and mission data of multiple aircraft are transmitted to ground stations where they are displayed in real time to the range training officer and recorded for replay during post mission debriefing. Tracks of the symbolized aircraft are presented on a three-dimensional (3-D) CRT display. The system also maximizes safety of training. If an aircraft violates any preprogrammed safety parameter, the pre-designated symbol for that aircraft begins flashing on the display screen. The ground instructor can then warn the pilot. Figure 1 portrays the four subsystems comprising TACTS.

---

[2] A new device, Air Combat Tactics Trainer, 2E7, is scheduled to be ready for training at NAS Lemoore, CA in late 1982, and a second unit for NAS Cecil Field, FL in mid-1983. Youngling et al. (1977) provide a description and evaluation of the several ACMSs used by the Air Force and NASA.

The first TACTS became operational in December 1973 at MCAS Yuma, AZ[3]. Ground stations for monitoring and debriefing ACM in the Yuma range are available at both Yuma and NAS Miramar. On the east coast, an over-water range became operational in 1976 off Cape Hatteras, NC. Ground training facilities are available at both NAS Oceana, and MCAS, Cherry Point, NC.

Existing TACTS Training Capabilities

The TACTS training capabilities provide: (1) highly realistic conditions approximating actual combat; (2) management, and direct control of training by ground-based instructors who can quickly vector participating aircraft into positions to set up for each engagement; (3) several missile envelope recognition training modes, varying progressively in difficulty; (4) immediate knowledge of results of simulated weapon release outcomes; (5) "nondebatable" (i.e., objective) magnetic tape record of certain actions in the aircraft; and (6) recorded flight and weapon action data for performance debriefing and hardcopy feedback.

Emerging TACTS Training Capabilities

The name change from ACMR to TACTS in 1980 reflects the broadening of anticipated training system capabilities beyond air combat maneuvering to a variety of aircrew combat tasks. Emerging capabilities include simulated air-to-ground weapons delivery (e.g., no-drop bomb scoring), electronic warfare (EW) training, and surface-to-air missile (SAM) avoidance training. Initial studies have been completed to investigate the feasibility of adding by FY86 at-sea monitoring and debriefing capabilities aboard aircraft carrier

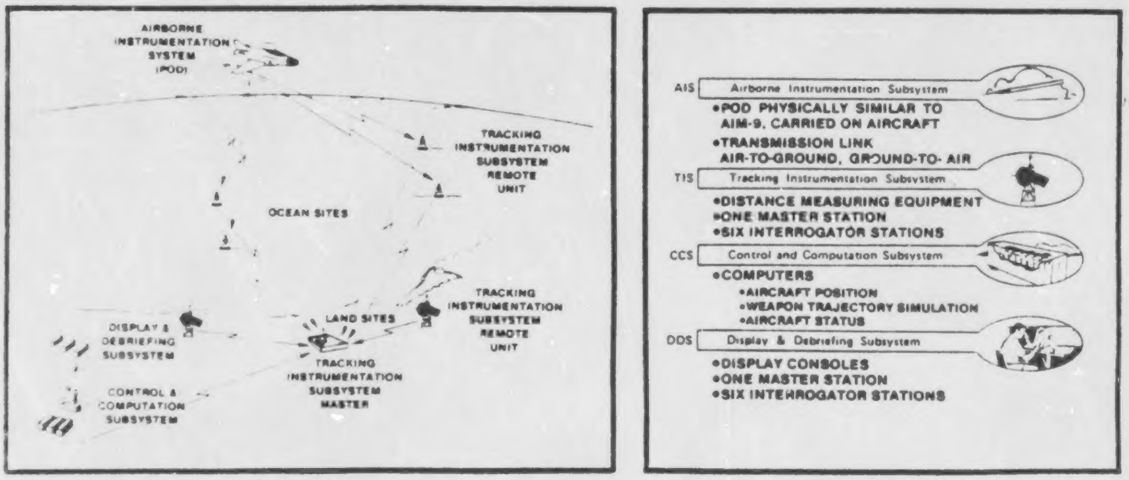

AIRBORNE INSTRUMENTATION SYSTEM (POD)
TRACKING INSTRUMENTATION SUBSYSTEM REMOTE UNIT
OCEAN SITES
LAND SITES
TRACKING INSTRUMENTATION SUBSYSTEM REMOTE UNIT
DISPLAY & DEBRIEFING SUBSYSTEM
TRACKING INSTRUMENTATION SUBSYSTEM MASTER
CONTROL & COMPUTATION SUBSYSTEM
AIS
Airborne Instrumentation Subsystem
•POD PHYSICALLY SIMILAR TO AIM-9. CARRIED ON AIRCRAFT
•TRANSMISSION LINK AIR-TO-GROUND, GROUND-TO-AIR
TIS
Tracking Instrumentation Subsystem
•DISTANCE MEASURING EQUIPMENT
•ONE MASTER STATION
•SIX INTERROGATOR STATIONS
CCS
Control and Computation Subsystem
•COMPUTERS
•AIRCRAFT POSITION
•WEAPON TRAJECTORY SIMULATION
•AIRCRAFT STATUS
DDS
Display & Debriefing Subsystem
•DISPLAY CONSOLES
•ONE MASTER STATION
•SIX INTEHROGATOR STATIONS

Figure 1. Description of the Four Subsystems of the Tactical Air Combat Training System (TACTS). Description is from the East Coast TACTS User's Guide (May 1978).

---

[3] The Air Force participated in the evaluation of the first Navy range and established a subsequent requirement for an Air Force range called the Air Combat Maneuvering Instrumentation (ACMI). The first ACMI became operational in July 1976 at Nellis AFB, NV, as a technologically updated version of the TACTS.

battle groups to the currently available land-based facilities (Crangle, 1980).[4]

DESCRIPTION OF ACM PERFORMANCE

The basic ACM task appears simple. The pilot attempts to: (1) destroy the enemy; and (2) avoid being destroyed by the enemy. According to John Johnson (1956), Britain's top fighter ace in WWII, pilots just before an ACM mission. . . "fall into two broad categories; those who are going to shoot and those who secretly and desperately know that they will be shot at . . ." He called the first group "hunters" and the second group "hunted".[5] Traditionally, however, the fighter pilot has been identified with the first, aggressive role and to a much lesser extent with the second, survival role. The cultural mystique surrounding the aggressive fighter pilot and the nature of his combat task are described below.

Task Complexity

During ACM the modern pilot must visually detect, identify, and track an adversary which can change position in 3-D geometric space at supersonic speeds. Meanwhile, the fighter must also rapidly manipulate his aircraft and weapon controls, monitor instruments and displays, and coordinate his actions over communication channels with other aircrew and/or friendly fighter aircraft. All these dynamic responses must take place within the typically short, 2-3 minutes, of an ACM engagement. In addition to the stringent perceptual-motor skill requirements, decision-making demands are continuous. For example, missile envelope recognition is a critical decision based on a number of factors, including the kind of aircraft adversary, fighter missile type selected, various rules of thumb for envelope determination, etc.

These perceptual-motor and decision-making skills are "enabled" only through the mastery of a third skill-control over emotional responses.[6] The pilot must keep his "cool" under conditions of high gravitational forces, the possibility of a ground or midair crash, and the extremely rapid loss of fuel during intense ACM on the TACTS. During actual combat, the pilot contends with additional stressors, including the likelihood of multiple adversaries with real weapons, communicaton jamming, SAM missiles and enemy "flak" of various kinds. Although 98 percent of pilots surveyed have highly favorable opinions of the instrumented ACM range as a training facility (Youngling et al., 1977), there was also consensus that the relative safety of the training environment limited to some degree transfer of training to the life-threatening episodes of combat ACM.

---

[4] Current ranges will also be modified as needed for operational test and evaluation (OT&E) research studies such as the Air Combat Evaluation - Air Intercept Missile Evaluation (ACEVAL-AIMVAL) program - a joint Air Force/Navy evaluation of new air intercept missiles and tactics.

[5] Weiss (1966) dichotomized combat pilots into "Hawks" (those that downed enemy aircraft) and "doves" (those that were shot down by enemy aircraft).

[6] A collaborative effort between the Naval Health Research Center/NTEC is planned for FY82 to identify training "stress profiles" of pilots during the experience of ACM and to relate these profiles to objective measures of ACM performance.

Air-to-air combat between aircraft is especially complex because in-close aircraft maneuvering is required. The classical and modern "dogfight" or "hassle" between a fighter and an adversary target (called a "bogey") takes place because of a requirement for visual identification (VID) of an adversary prior to weapon release. The operational requirement for VID lessens the likelihood of destroying friendly aircraft. However, VID is not necessary for existing missile effectiveness which can extend well beyond visual acquisition range, and beyond the airspace in which visually controlled dogfighting would otherwise occur.[7]

The abbreviation "ACM" is often used as a catch-all expression for the entire air-to-air combat mission which is highlighted by the aircraft maneuvering engagement, (dogfight). The entire mission includes several kinds of performance skills, including radar procedures, VID intercept procedures, tactics, ACM, weapon system and missile/cannon envelope recognition, and "bugout" (return to base) procedures.

The "Fighter Mystique"

Until recently, attempts to quantify the highly complex ACM task and to reduce it to elemental sequences of subskills were scarce. The ACM performance measurement (ACMPM) methods described below are all recent products of the last decade. Part of the historical reluctance to develop ACMPM systems can be attributed to what might be called the "fighter mystique" (Eddowes, 1981). There appears to be a general sentiment in the populace that fighter pilots have the "right stuff" (Wolfe, 1980). It is also generally acknowledged that the ingredients for the right stuff are complex and unknown. Even if the ingredients were known, they could not be quantified. Finally, the argument continues, even if they could be quantified, so few people would have the right stuff that measurement would be impractical.

Studies of combat reviewed by Youngling et al. (1977) support the general view that only a few fighter pilots have the right ingredients for combat success. For example, only five percent of the 5,000 fighter pilots who flew against the Germans in WWII during 1943-1945 accounted for 40 percent of the kills recorded by the Eighth Air Force during that period. Even more impressive were the German fighters whose ten best chalked up 2,568 kills among them (Weiss, 1966).

The unfortunate consequence of the fighter mystique attitude present in this culture is that the skill components of the ACM task have not been translated into performance measures which could be used to provide training feedback for the majority of fighter pilots who could profit from such feedback. Data from WWI, WWII and the Korean conflict indicate that less than 15% of all fighter pilots had a better than even chance of surviving their first combat experience (Weiss, 1966). Based on data such as these, it has been estimated that, without specialized air-to-air combat training, high attrition rates can be expected in future conflicts (Youngling et al., 1977).

---

[7] Certain navy combat experienced ACM pilots (e.g., Flynn, 1975) feel that the advent of new missile technology "will not eliminate the need for visual identification of enemy aircraft and hence the inevitable dogfight" (p. 4). An identical view has been expressed by Air Force pilots (Ethell, 1980).

EXISTING TACTS ACM PERFORMANCE MEASUREMENT (ACMPM) CAPABILITIES AND LIMITATIONS

As described above, TACTS contributes directly to training by providing an ACM performance measurement capability. TACTS records and plays back objective performance data in a variety of graphic and numeric formats. However, the existing TACTS ACMPM has a number of specific limitations for training which will be described below. The full potential of ACMPM has yet to be realized. As evidence of the need for a system of performance measurement on the TACTS, the Navy Fleet Project Team (FPT) for air combat training, recently stated:

> "A performance measurement system (PMS) is required to determine training effectiveness, act as a diagnostic training aid, and evaluate present/future training device capabilities relative to the syllabus. As a basis for PMS, indices must be identified. A standard PMS should be established." (COMTACWINGSLANT naval message, March, 1981.)

Limitations in Existing ACMPM

The limitations indicated below confirm the conclusions of the FPT and provide a foundation for the development of a set of requirements for an ACMPMS in support of the fleet.

Data Flooding. The performance measures available during ACM and for the ACM debrief are so "rich" that some users have described the scheme as "data flooding". The instructor is genuinely overloaded with such an array of information.

Lack of Specific Training Objectives. Data flooding is a natural consequence of inadequate specification of behavioral objectives. When it is uncertain as to what is being trained, the simple (and often more costly) solution is to measure everything.

Lack of Trend Data. A major limitation of the TACTS is that performance data are not accumulated over engagements to identify aircrew trends (Ciavarelli, Williams, & Stoffer, 1981). Trend data are necessary to identify and to diagnose consistent patterns of performance as a baseline for corrective action.

Certain Important Variables Unmeasured. A second consequence of inadequate specification of training objectives is that an important performance variable may go unmeasured. One such variable, time in envelope, is not available under the present TACTS PM. This omission is interesting since envelope recognition skill training was one of the three original requirements leading to development of TACTS.[8] A second variable, energy management, has only recently been available as a debriefing aid at NAS Miramar. An energy management display (EMD) measures the maneuvering capability of the aircraft by integrating energy maneuverability data available from the TACTS. The ability to maximize maneuvering capability is a major determinant of the outcome of air combat (Deberg, 1977; Pruitt, Moroney, & Lau, 1980).

---

[8] Even if a time in envelope indication were immediately available, it would probably be necessary to reconstruct or qualify the variable (i.e., on the basis of a rate measure).

Inadequate Debrief Data Formatting. Much of the performance data available during the debrief is in numeric form which is difficult to visually process and retain. The operational acceptance of the EMD is based in large part on the fact that it is a pictorial display of the energy maneuvering envelope of an aircraft and of its opponent during an ACM engagement. The EMD was developed as a debriefing tool with considerable input from TACTS pilots. Similarly, the complexity of missile envelopes would be more easily understood if represented by graphic than numeric formats. An ongoing project, directed by the Human Factors Laboratory (HFL) at the Naval Training Equipment Center, has focused on the development of methods to standardize the entire debrief feedback in terms of display tables and graphs that are recommended by TACTS users (Ciavarelli, Pettigrew, & Brictson, 1981). Methods for missile envelope representation, and simultaneous representation of associated missile shots, are now available.

Lack of Quality Control Over Raw Performance Data. Two major emerging systems to provide performance measurement for TACTS will be described in a subsequent section. Both of these efforts, as well as a more recent effort with a similar goal (McGuinness et al., 1980), have had to implement considerable quality control over the performance data prior to data analysis and interpretation. McGuinness et al. have described in some detail a number of limitations in TACTS hardware and operational procedures which consequently require "filtering" of TACTS output data.

VALUE OF TACTS AS A SOURCE OF PERFORMANCE CRITERIA

The objective performance data available from TACTS are sources of valuable criteria for several purposes, including training.

Selection

One of the original (1975) research interests in the TACTS ACMPM was to identify, define and validate behavioral criterion variables for use by the Naval Aerospace Medical Research Laboratory (Brictson, Ciavarelli & Jones, 1977; Brictson et al., 1978). The Vision Research Laboratory, in particular, was interested in developing realistic criteria for validation of certain visual variables used to select naval aviators. There appeared to be an obvious mismatch between current selection variables which emphasize static, high contrast, central vision acuity and the dynamic, low contrast, peripheral vision characteristic of ACM - a major mission of the navy fighter pilot community. In a study using TACTS data, e.g., it was found that the average visual target acquisition ("tally-ho") range was considerably shorter than expectation based on laboratory data on human visual capability (Hutchins, 1978). These laboratory data are the basis for the visual selection tests currently in use by the Navy. Youngling et al. (1977) have systematically reviewed and recommended ACM ranges as intermediate test beds for combat effectiveness predicted by various selection factors.

Training

Performance data are essential to the training process. Measures of ACM subtask skills allow the instructor to monitor the progress of training, as well as to provide diagnostic feedback regarding problem areas. Both norm-referenced and criterion-referenced standards of aircrew proficiency can be established. These behavioral criteria and associated standards inform the instructor and student as to what is to be trained and to what level. Training effectiveness evaluations (TEEs) are then possible.

Assessment

Assessment of operational ACM readiness, tactics, equipment, safety, and cost-effectiveness of the range, is possible only with the availability of TACTS performance data.

Readiness Assessment. Once combat becomes necessary, the outcome rests on the quality of the previous aviator selection process and training in ACM. A third factor is placement or deployment of aircrew and squadrons on missions that are suited to their level of operational readiness. Readiness assessment is not concerned with changing the level of ACM proficiency through training; it simply attempts to describe existing squadron and individual differences among fleet fighters.

Tactics Assessment. Performance data taken during TACTS engagements can update ACM tactics. The Naval Fighter Weapons School (NFWS), called "Topgun," is responsible for teaching advanced ACM tactics. Lessons learned from tactics assessment on the TACTS have already begun to update Topgun tactical guidance, including those that concern throttle control during ACM and various "rules of thumb" for recognizing missile launch envelopes.

Equipment Assessment. The technology of TACTS is changing rapidly with the development and introduction of new aircraft (i.e., the F/A-18), new air intercept missiles (i.e., all aspects capable), new performance aids incorporated in Heads-Up-Display (HUD) to improve missile envelope recognition (Lutter, 1979), new EMD performance aids developed to maximize the maneuvering capability of the fighter aircraft (Pruitt et al., 1980), etc. Performance criteria will serve to evaluate whether the use of modified capabilities and performance aids translates into improved performance on TACTS.

Safety Assessment. Similar to commercial airline recording of flight data and human performance for subsequent analysis for safety factors, TACTS allows both on-line and off-line capability for safety assessments leading to corrective and/or preventive feedback.

Cost-Effectiveness Assessment. The use of simulated weapons and targets has greatly decreased the cost (and increased the safety) of ACM practice. The producer of the instrumented air combat ranges for both the Navy and the Air Force estimates that the TACTS/ACMI reduces air combat training costs by more than $100 million annually (Cubic Corporation, 1978). According to Bill Dollard (1980), the TACTS manager at Miramar, the first system developed for Miramar cost approximately $25 million in R&D and installation of the system in an operational mode. The system costs less than a $1,000 an hour to operate. However, these savings must be assessed in terms of performance effectiveness measures on the TACTS.

Simulation Design

One of the most important values of PM is its role in training device design. Future Navy air combat training simulators will be designed with the benefit of TACTS performance data that has validated certain ACM simulator training capabilities and not others. In particular, the simulator's PMS can be designed to overlap that available on the TACTS. Common performance measures between ACM trainers and TACTS would provide an ideal foundation for effectiveness evaluations of simulators (see McGuinness et al., 1980).

MAJOR APPROACHES TO ACM PERFORMANCE MEASUREMENT (ACMPM)

The literature on ACMPM can be organized on the basis of common approaches to the problem of measurement.

Systems View vs. Component View

Most researchers in this field have taken a systems view of ACMPM. Thus, human performance and its measurement is viewed as only one component dependent on other elements in an ACM training system. The system includes such elements as aircrew, instructors, aircraft, weapons, and operating conditions, all of which impact performance output.

Multivariate vs. Univariate Analyses

Most researchers have recognized the complexity of ACM success and have adopted various statistical techniques to deal with this complexity. Not surprisingly, multivariate methods such as multiple regression (Coward et al., 1979), factor analysis (Deberg, 1977), and discriminant function analysis (Kelly et al., 1979), have been applied primarily with ACM simulators rather than with the TACTS. In contrast to TACTS, these simulators allow sufficient pilot sample size and repeated measures on selected variables to allow the use of multivariate statistical approaches to the development of ACMPM.

The Readiness Estimation System (RES), described below, provides a performance index of ACM based on complex mathematical models that incorporate a number of airplane and inter-airplane parameters typically measured by the TACTS.

It should be noted that operational users (Seminar, 1980) of TACTS are generally opposed to a simple, univariate characterization of performance. Their legitimate fear is that ACM training would be geared to achievement of that particular performance metric to the exclusion of others critical for ACM proficiency. In addition, there appears to be a desire to avoid labeling the performance of any particular pilot based on one measure that might have adverse effects on his competitive spirit. Resolution of the apparent conflict between research development of single metrics and the concerns identified above will be found, in part, by development of a composite measure of ACM success, dependent on multiple tasks, with each task weighted according to its relative contribution to overall ACM success. This composite measure, under development by the Human Factors Laboratory, will allow both subtask and composite ACM task performance feedback, to supplement the traditional measure of ACM success in terms of fighter/adversary kill exchange ratios.

"Building Blocks" vs. "Top-Down" View of ACM Success.

Some (e.g., Ciavarelli, Williams, & Krasovec, 1980) have chosen a hierarchical or "building blocks" system approach to development of a PMS by first focusing on specific ACM tasks that are evaluated at various layers in the performance hierarchy for criticality to ACM success. Others use a "top-down" approach by focusing on ultimate or penultimate performance measures of ACM success and then proceeding to subtask analysis. For example, McGuinness et al. (1980) have sharply criticized approaches based strictly on subtask analysis. At some level, however, agreement between these two approaches will eventually be reached.

Performance variables at different levels in the measurement hierarchy are useful for different purposes. For example, task and subtask approaches provide sufficient detail for individual pilot diagnostic feedback. At this level it is possible to find out what specific skills superior fighter pilots possess so that these skills can be trained in new pilots. Approaches that focus on system output level such as exchange ratios provide valuable information for overall squadron readiness assessment. The RES, e.g., provides overall ranking of aircrew and squadrons on the basis of overall maneuvering scores, but does not allow for individual diagnostic information to explain differences in maneuvering proficiency. At the system level are output variables that are used to assess the relative contribution of various subtasks and to construct a composite measure of ACM success referred to above. Thus, not only do performance measures at various levels in the hierarchy serve useful purposes by themselves but they also complement each other in a complete system of performance measurement.

Dynamic vs. Static Representations of Performance

The original TACTS provides qualitative feedback through dynamic replay of the entire engagement. Some more recent PM schemes (i.e., the RES) provide performance feedback regarding the dynamic interplay between opposing aircraft. The RES provides a time-history, quantitative index score of maneuvering performance based on continuous measurement throughout the engagement.

Other approaches conceptualize maneuvering as one rather discrete step in a total ACM mission sequence preceded by radar procedures and lookout procedures, and followed by envelope recognition procedures (Ciavarelli et al., 1981). Obviously, the dynamic properties within each of these discrete "steps" could be elaborated into a PM metric. For training diagnostic purposes, however, there is considerable value in providing initial performance feedback in the form of discrete and comprehensible steps involved in the task to be trained. The more complex and less comprehensible performance dynamics associated with each step should be reserved until the pilot has achieved minimal proficiency at those particular steps. In reality, the distinction between dynamic and static representations of performance becomes blurred to the extent that sequential dependencies between discrete events in the ACM mission can be established, as has been demonstrated by Ciavarelli and his associates.

Preselected Pilots vs. Preselected Variables

Most research in ACMPM falls into one of two approaches. One approach (e.g., Kelly et al., 1979) first preselects pilots high or low in overall ACM proficiency, on the basis of instructor ratings, simulator performance, or past TACTS records. These two groups are then contrasted on the basis of a large number of specific performance measures. Measures that differentiate the two groups are prime candidates for the PM scheme. The second, and more common approach, preselects a limited number of specific variables for possible inclusion into the PM scheme and then tests them for relative reliability and validity across a wide range of pilot proficiency levels.

Air Combat Maneuvering Simulator (ACMS) vs. TACTS

Ideally, the design of ACM simulators and their associated PM system is preceded by development of a PM system validated on the TACTS range itself. Unfortunately, none of the existing Navy ACMSs have benefitted from a TACTS PMS. Thus, ACM simulators have been recently used not only to develop ACM PMSs for use with the simulators themselves, but also as a major approach for developing a possible PMS for TACTS. Through an iterative process, there will eventually be developed a set of measures common to the ACMS and TACTS.

In addition to the obvious benefits of increased safety and economy in an opertaional training area traditionally characterized by high risk and high cost, ACMSs allow pilots to experience a wide variety of tactical situations and gain more practice with these situations than is currently possible on the TACTS. Thus, with the ACMS, it is feasible to actively control training experience and to achieve repeated measures of performance data - both of which are essential for PMS development and extremely difficult with the TACTS. The only systematic attempt to develop an ACMS PMS with the Navy is just beginning, with the 2E6 at Oceana, under the direction of the HFL.[9]

FUNCTIONAL REQUIREMENTS FOR AN ACMPM SYSTEM

In view of the above described limitations of original TACTS PM, the recent FPT message stating the need to develop a PM system for use with the TACTS, and the tremendous potential value of TACTS performance criteria, it is clear that functional requirements for a PM system should be identified. As indicated previously, a system of ACM measurement is yet to be developed.

The purpose of this section is to identify several psychometric, training, and user feasibility requirements that must be met for a PMS to be developed. These functional requirements will be briefly described using two emerging TACTS ACMPMSs. Both systems use the original TACTS ACMPM as a basis for elaboration. Both emphasize highly automated methods to manage the TACTS outputs. It is not intended to comprehensively compare the two or to recommend either one or the other PMS. If anything, both systems are recommended for further development. They were chosen to illustrate PMS requirements because they are the only systems that: (1) have been under development for several years; (2) meet many of the requirements identified below; (3) have a considerable body of technical documentation available for evaluation; (4) are familiar to operational users on both coasts; and (5) have been accepted or implemented to some degree on both coasts.[10] These two PMSs will be briefly described below.

The Readiness Estimation System (RES) is an automated, off-line and time-based measurement system under development that continuously calculates the relative positional advantage/disadvantage of the fighter aircraft at a specific point in the engagement. It provides useful data for envelope recognition plus tactics and maneuvers. Plotted data provide information related to both time history, and time in envelope which is not available on the original TACTS. The RES was developed by Simpson (1976) and Simpson and Oberle (1977).[11]

---

[9] There are two other emerging and continuing Air Force ACMS PMSs - one called Tactical Space (TACSPACE) under joint AFHRL/HFL sponsorship (Kelly et al., 1979), and one called the Good Stick Index (GSI) developed by Coward et al. (1979).

[10] The Fleet Fighter ACM Readiness Program (FFARP) generates the "Blue Baron" reports that summarize individual squadron performance data taken directly from TACTS output. The FFARP is a relatively structured and concentrated three-week syllabus of ACM sorties on the TACTS. It is conducted annually by an adversary squadron. FFARP exercises, until recently, were limited to east coast TACTS. The FFARP analyses are regarded herein as an incomplete system of PM, primarily because they contain most of the limitations of the original on-line TACTS ACMPM. They do provide valuable hardcopy record of data manually retrieved from the TACTS. However, squadron feedback from FFARP exercises is currently taking in excess of two weeks and the performance results are costly to generate.

The Performance Assessment and Appraisal System (PAAS) is an automated off-line system under development to provide a systematic ACM performance debrief, including computer graphic formats, multiple-referenced standards of performance, diagnostic analyses of ACM subtask, and squadron performance summaries. It can be described as an information management system. PAAS is presented in more detail by its developers in a separate paper in these proceedings (Ciavarelli, Williams, & Brictson, 1981).

Psychometric Requirements

Psychometric requirements are the traditional and necessary ones that establish the statistical soundness of a measurement system. They represent an application of the more general requirements for any measurement system to human performance measurement. Several specific requirements are identified below.

Objectivity. Objective raw data indicators of performance, routinely available from TACTS, are the foundation for the two emerging TACTS PMSs. Without objectivity, there is no system of measurement because everyone is free to construct his own subjective "system". Since the TACTS can provide objective criteria, long lacking in both selection and training of military aviators, North and Griffin (1977) refer to the possibility of TACTS as an "ultimate" criterion for evaluation of naval aviator selection variables in a highly realistic, advanced training environment. Similarly, since scientists are no longer limited to instructor or peer ratings to assess training, some researchers (Ciavarelli & Brictson, 1978) have referred to the TACTS PM capabilities as the dawning of a "golden age" in operational training PM.

Reliability. Objective measures must be reliable. Reliable performance measures provide the consistency and permanence necessary for system stability. Although there are several methods for measuring reliability, one most appropriate for ACM training is temporal reliability. If, e.g., enemy aircraft visual identification (VID) accuracy is a reliable measure, a particular pilot should be equally proficient (or non-proficient) at the VID task across several recordings of VID performance. Thus, reliability is the basis for the diagnosis of consistent pilot trends in performance. Diagnosis of individual pilot performance allows corrective action of substandard performance through individualized training. Unfortunately, neither the RES nor the PASS have documented the reliability of the measures in their PMS. The major limitation to reliability determinations has been the typically small number of repeated measures of performance available for specific aircrews using the TACTS.

Validity. System validity exists only if the measurement variables are meaningful. VID performance, e.g., gains its meaning only through its correlation with other variables, regardless of how consistent or reliable VID performance might be. VID correlates with air combat engagement success, both according to tactical doctrine (NFWS) and empirical research tests of actual performance on the TACTS (Brictson et al., 1977; Ciavarelli, Brictson, & Young, 1979). Thus, the VID subtask measure has predictive validity as it can predict ACM success.

---

[11] A major modification of the RES, called the Readiness Index Factor (RIF) has been made by operational aircrew members at the Oceana TACTS. The RIF is currently under evaluation by the HFL as a performance measurement tool for use in a study examining the transfer of training from an ACM simulator (2E6) to the TACTS.

Concurrent validity is present when one performance measure correlates with other performance measured concurrently (or nearly concurrently) in time. For example, a fighter pilot is twice as likely to get a VID on the TACTS if he has gotten radar contact than if he has not (Ciavarelli et al., 1980). Confirmation of such sequential probabilities between ACM subtasks is also in agreement with tactical doctrine.

An ACMPMS is not completely valid without content validity, which exists when the performance measures are inclusive or at least representative of the ACM task. Thus, a thorough job analysis of ACM is required to ensure that the performance measures selected represent the kinds of subtask content required for ACM engagements. Although the RES has very little predictive or concurrent validity documentation available, it is strong in content validity because it focuses its measurement system on the maneuvering task (ACM) which is uniformly regarded as the most visible and dramatic content of air-to-air combat.

Sensitivity. A measurement subsystem is sensitive to other subsystems of the overall training system. If there is a PMS, it is, by definition of a system, interrelated with other subsystems. A system of ACM performance measures, therefore, will be sensitive to changes in other TACTS subsystems such as aircrew, instructors, weapons, etc. To illustrate, the PAAS measures are highly sensitive to differences in commonly acknowledged aircrew proficiency levels between fleet operational squadrons and the typically more experienced fleet reserve fighter pilot squadrons (e.g., Ciavarelli et al., 1980). Similarly, differences in thrust to weight ratios and wing loadings of different aircraft are routinely reflected in the RES metric.

Quality Control. A system of measurement does not admit as input data that does not meet certain quality of raw data requirements. Only through systematic filtering of input does a system maintain its characteristic and recognizable features. Moreover, such quality control is necessary to achieve reliability, validity, and sensitivity of the PMS. Both the PAAS and the RES have quality control procedures for data entry into the PMS.

Training Requirements

Training requirements for a PMS are those specifically appropriate for use in training applications. They are met most completely when the above psychometric requirements have been met.

Descriptive. A system of measurement should, at minimum, provide a complete description of relevant performance. The RES, e.g., traces continuously the maneuvering skills of the pilot over the course of an engagement.

Evaluative. In order to standardize training, it is necessary to evaluate performance in terms of established standards of performance. Common standards against which individual performance may be gauged are a necessary part of a PMS. At least two kinds of standards are required to fully evaluate ACM performance. Criterion-referenced standards are based on an absolute reference point that represents the criterion performance to be reached by aircrew members. For example, tactical doctrine prescribes missile envelope boundaries outside of which missile launches would be classified as errors. It is possible to compare directly tactical envelopes (the criterion) with actual shots (placements) taken in TACTS to verify that shots were within the prescribed boundaries.

Since all pilots do not always meet criterion on all ACM subtasks, it becomes necessary to describe the variability of performance on different subtasks. Norm-referenced standards describe the average or normative performance, regardless of preestablished criterion levels. Normative performance is described essentially by using group averages with a second statistic representing variability of individual performance around that average. For example, tactical missile envelope boundaries can be compared to normative envelopes based on empirical data taken from the TACTS. One practical use of normative standards is to develop criterion standards from them. DeLeon (1977), e.g., suggests that training to gain incremental increases in air-to-air combat skills above the average skill levels is more reasonable than trying to train everyone to be an ace. Both kinds of standards have been developed for some subtasks (cf., Ciavarelli et al., 1980). Ideally, users will soon be able, via automated PMS features, to select and create their own performance norms for various TACTS operating conditions.

Diagnostic. A PMS provides the opportunity for self-correction on the basis of performance feedback. The system provides enough detail in its diagnosis of tasks or subtasks that the individual can recognize those specific behaviors that are not currently meeting established standards. The RES was not constructed to be diagnostic of individual aircrew tasks but to represent the overall level of aircrew readiness. The diagnostic capability for PAAS is just now beginning to emerge.

Remedial. A PMS not only informs an individual of his skills (i.e., diagnosis), and the level of skill achievement (i.e., evaluation), it also identifies specific corrective actions in the event that remedial action is necessary. Neither PASS nor RES provides remedial training currently.

Timely. To maximize the self-corrective feature of a PMS, feedback should be as rapid and as continuous as possible. Because the PASS and RES are off-line systems that cannot operate while TACTS training is in progress, neither PMS provides immediate ACM feedback after an ACM engagement. However, both systems can provide extensive feedback within 24-48 hours.

User Feasibility Requirements

User feasibility requirements are essential to a PMS because if they are not met the "system" will not be used. These requirements basically deal with the operational user's cost, ease, and overall acceptability of operating the PMS. They are even more specific than the training requirements illustrated above because they apply to ACM training in particular and the feasibility issues associated with that specific application.

Affordable. At this point, both the PAAS, and particularly the RES, require an analyst to process and interpret the performance feedback from the PMS. Before either PMS gains long-lasting use, it will be necessary to minimize those costs which support the system.

Adaptable. Changes induced by system growth, and changes in user needs, including changes in training, must not disrupt the basic operation of the

PMS. Design modularity is essential. Thus, it should not be essential to continue measuring out-of-date variables in order to add a new and necessary measurement variable. Likewise, it should not be essential to eliminate still useful variables in order to drop an old and now unnecessary variable. The PASS, and to a lesser extent, the RES, are modularized so that they will not lose their system status due to system growth.

Automated. Much of the data currently available to users of the TACTS are only available with a great deal of manual extraction efforts. Since instructor and student workload can prevent use of otherwise valuable performance feedback, both the RES and PASS are automated.

Manageable. Designers of the PMS should reduce the number of feedback variables to those required for meeting behavioral training objectives. The data flooding referred to earlier is evidence of the confusion that a PM system serves to reduce. The PASS and RES share a significant advantage in this respect.

Accessible. For performance data to be used, they must be easily accessible by the user subsystem that interacts with the PMS subsystem. An ideal PMS makes it rather simple for users to both enter and retrieve performance data. The PAAS, in contrast to the RES, was designed especially with accessibility in mind, by incorporating a number of "user friendly" computer features.

Non-Interference with Training. Both PMSs are currently off-line because their on-line operation would require dedicated TACTS computer time which would decrease the amount of already scarce ACM training time. Although the ultimate goal is to gain on-line PMS capabilities, off-line processing is the only currently feasible alternative.

Acceptable. Acceptability is an umbrella term for all user requirements that determine whether TACTS users will, in fact, be motivated to use the PMS. For example, the RES, a PMS with a great deal of technical merit, has undergone major modifications at Oceana, partly because it was not easily understood or used. Without user acceptability, even a well-validated and low cost PMS has no value. Thus, certain ACMPMS features that have more to do with motivational than instructional capabilities per se are required for acceptability. Motivational factors that impact system utilization include user choice of alternative graphic display formatting of performance feedback, possibility for instructor override and flexible use of automated PMS, administrative record keeping assistance to the instructor by the PMS, privacy coding of performance results, etc.

CONCLUSIONS: TACTS TECHNOLOGY AND HUMAN PERFORMANCE MEASUREMENT

Without question, the TACTS is an impressive application of state-of-the-art technology in airborne instrumentation and tracking, computer-generated simulation of weapon launch outcomes, video tape playback of voice transmissions and graphic portrayal of flight history data. Equally impressive, however, is the conclusion that the TACTS was developed and is currently operated without the benefit of a system of performance measurement. This conclusion is based on several observations: (1) identification of several limitations in the TACTS capabilities for training; (2) a recent (March 1981) naval message from the FPT for air combat training that an ACMPMS was needed; (3) recent attempts by operational users of the east coast TACTS facility to develop their own PMS; and (4) a mismatch between the original TACTS PM scheme and a set of psychometric, training, and user feasibility requirements for a PM system (PMS).

In terms of the three original requirements leading to development of TACTS, the third requirement, for safety and economy of training, has been achieved. Fulfillment of the second requirement, for a performance debrief capability, depended in large part on the development of a PMS and, therefore, has been only partially achieved. The first requirement, for missile envelope recognition training, is difficult to evaluate because there is not one envelope, but multiple envelopes to be learned for different missiles and aircraft capabilities. In addition, there is no completely uniform agreement as to the "proper" firing envelope for a specific missile. Thus, it is not surprising that several systematic samples of launch data obtained on the TACTS indicated that even experienced pilots often fire outside of prescribed doctrinal missile envelope limits (Ciavarelli, Narsete & Brictson, 1981).[12] Such variability in performance is in part due to the lack of a PMS for standardization of envelope recognition training.[13]

The theme for this symposium was, "Is advancing technology ignoring human performance in aviation systems?" It would appear that with the development of the TACTS instrumentation technology, there was no parallel development to provide a PMS to support that capability. This apparent "oversight" is a little easier to understand when one considers the almost mind-boggling complexity of the ACM task and the fighter mystique which is the cultural heritage associated with ACM. In addition, during production of the original TACTS, there was not available a well-developed PMS that could have been further developed.

It was suggested that there are several viable approaches to the measurement of TACTS performance. Two promising and emerging systems of ACMPM were described and used to illustrate a set of functional requirements for a PMS. Fulfillment of these requirements will provide valuable performance criteria for a number of different purposes, including maximization of the training capabilities of the TACTS, and will allow the "golden age" of operational performance measurement to mature.

---

[12] A major advantage of TACTS is that it allows such systematic evaluations of envelope recognition skills.

[13] A second major reason is that TACTS users do not generally use the five available envelope recognition training modes which vary progressively in level of difficulty and amount of feedback provided. The strong preference is to use mode five which is the most realistic and challenging one. A third significant reason is that the TACTS provides very little envelope recognition practice for any particular aircrew. Access to the range is highly competitive. During TACTS, the engagements are usually no more than several for each aircrew and engagements typically last less than a few minutes. Users (e.g., Carter, 1979) point out that the recognition skill is "achieved only through repetitive exposure to possible firing situations (p. 68)". An analytical effort prior to TACTS development to identify recognition training capabilities would almost certainly have concluded that simulator training would be needed to supplement practice on the TACTS.

REFERENCES

Applied Physics Laboratory, Johns Hopkins University. Final report on ACMR evaluation: Data collection instrumentation operational suitability, Volume II, Contract 600084. Flight Systems Inc., May 1975.

Ault Committee Report. Report of the air-to-air missile systems capability review. COMNAVAIRSYSCOM, January 1969.

Brictson, C.A., Ciavarelli, A.P., & Jones, T.N. Development of aircrew performance measures for the air combat maneuvering range (ACMR) (U). Naval Aerospace Medical Research Laboratory, Technical Report L53001, June 1977. (CONFIDENTIAL)

Brictson, C.A., Ciavarelli, A.P., Pettigrew, K.W., & Young, P.A. Performance assessment methods and criteria for the air combat maneuvering range (ACMR): Missile envelope recognition (U). Naval Aerospace Medical Research Laboratory, Special report 78-4, July 1978. (CONFIDENTIAL)

Carter, L., II. Fighter crews: You fight the way you practice. Proceedings of the Naval Institute, 1979, April, 65-69.

Ciavarelli, A.P., & Brictson, C.A. Air combat maneuvering range: Has operational performance measurement entered a golden age? Human Factors Society, Detroit, MI, October 1978.

Ciavarelli, A.P., Brictson, C.A., & Young, P.A. Development and application of performance criteria and aircrew assessment methods for the air combat maneuvering range (ACMR) (U). Naval Aerospace Medical Research Laboratory, Special Report 79-5, September 1979. (CONFIDENTIAL)

Ciavarelli, A.P., Narsete, E.M., & Brictson, C.A. A problem definition study of U.S. air-to-air missile envelope recognition training. Naval Training Equipment Center, Technical Report 78-C-0136-2, June 1981. (Confidential)

Ciavarelli, A.P., Pettigrew, K.W., & Brictson, C.A. Development of a computer based air combat maneuvering range debrief system. Naval Training Equipment Center, Technical Report 78-C-0136-1, June 1981.

Ciavarelli, A.P., Williams, A.M. & Brictson, C.A. Development and application of air combat performance assessment methods. Symposium on Aviation Psychology, Columbus, OH, April 1981.

Ciavarelli, A.P., Williams, A.M., & Krasovec, K. Operational performance measures for air combat: Development and application. Human Factors Society, Los Angeles, CA, October 1980.

Ciavarelli, A.P., Williams, A.M., & Stoffer, G.R. Training improvements for the tactical aircrew training system (TACTS): Project summary report. Naval Training Equipment Center, Technical Report 78-C-0136-3, June 1981.

COMTACWINGSLANT, Oceana, VA. Naval Message 232030Z March 1981, Subj: Report of Meeting of Fleet Project Team (FPT) for Air combat Training. (U)

Coward, R.E., Moore, S.B., Madison, W.G., Sepp, G.D., & Stracener, J.T. Air combat training: Good stick index validation. Air Force Human Resources Laboratory, AFHRL-TR-79-15, June 1979.

Crangle, R.H. TACTS program review, Naval Air Systems Command (AIR-06EA). Proceedings of the Seventh TACTS/ACMI User Seminar, San Diego, CA, May 1980.

Craven, P.R. The TACTS/ACMI Microcosm. Sea Power, 1980, December, 41-42.

Cubic Corporation. Cubic corporation capabilities, October 1978.

DeBerg, O.H. A scoring algorithm for assessing air-to-air combat performance, phase I: A factor analysis. Air Command and Staff College, Air University, Report 0590-77, May 1977.

DeLeon, P. The peacetime evaluation of the pilot skill factor in air-to-air combat. The Rand Corporation, R-2070-PR, January 1977.

Dollard, W. Cited by K. Cooper in, Navy uses 'TV game' replays to improve air combat training. San Diego Union, September 14, 1980.

Eddowes, E. Measuring pilot air combat maneuvering performance. Symposium on Aviation Psychology, Columbus, OH, April 1981.

Editors. You fight like you train, and TOPGUN crews train hard. Armed Forces Journal International, 1974, May, Pp. 25-26, 34.

Ethell, J. The modern dogfight. Air Progress, 1980, 42 (June), Pp. 41, 66, 68.

Flynn, S.C., Jr. Flight simulation and the "dogfight". Unpublished manuscript, 1975.

Futrell, R. The United States Air Force in Korea, 1950-1953. New York: Duell, Sloan, & Pearce, 1961.

Hutchins, C.W., Jr. The relationship between air combat maneuvering range (ACMR) output measures and initial visual acquisition performance. Naval Aerospace Medical Research Laboratory, Special Report 79-1, August 1978.

Johnson, J.E. Wing leader, London: Chatto & Windus, 1956.

Kelly, M.J., Woolridge, L., Hennessy, R.T., Vreuls, D., Barnebey, S.F., Cotton, J.C., & Reed, J.C. Air combat maneuvering performance measurement. Air Force Human Resources Laboratory, Flying Training Division, AFHRL-TR-79-3, September 1979.

Lutter, R.N. Information transfer for improved pilot performance. AGARD Aerospace Medical Panel Specialist Meeting, Lisbon, Portugal, October 1979.

North, R.A., & Griffin, G.E. Aviator selection 1919-1977. Naval Aerospace Medical Research Laboratory, Special Report 77-2, October 1977.

McGuinness, J., Forbes, J.M., & Rhoads, J.E. Applied training effectiveness evaluation methodology. *Naval Training Equipment Center*, Technical Report 79-C-0016-1, April 1980.

Pruitt, V.R., Moroney, W.F., & Lau, C. Energy maneuverability display for the air combat maneuvering range/tactical training system (ACMR/TACTS). *Naval Aerospace Medical Research Laboratory*, Special Report 80-4, October 1980.

Seminar, Aircrew Performance and Readiness Evaluation. NAS Oceana, VA, March 1980.

Simpson, W.R. Development of a time invariant figure-of-merit for use in analysis of air combat maneuvering engagements. *Naval Air Test Center*, TM 76-1 SA, July 1976.

Simpson, W.R. & Oberle, R.A. The numerical analysis of air combat engagements dominated by maneuvering performance. *Naval Air Test Center*, TM 77-2 SA, June 1977.

User's guide: Air combat maneuvering range (east coast). SRI International, COMNAVAIRSYSCOM No. 78-C-0090, May 1978.

Weiss, H.K. Systems analysis of limited war. *Annals of Reliability & Maintainability*, 1966, 5.

Wolfe, T. *The right stuff*. New York: Bantam Books, 1980.

Youngling, E.W., Levine, S.H., Mocharnuk, J.B., & Weston, L.M. Feasibility study to predict combat effectiveness for selected military roles: Fighter pilot effectiveness. *Defense Advanced Research Projects Agency*, Report MDC-E-1634, April 1977.

DEVELOPMENT AND APPLICATION OF AIR COMBAT PERFORMANCE ASSESSMENT METHODS

A. P. CIAVARELLI, A. M. WILLIAMS, and C. A. BRICTSON

Dunlap and Associates, Inc.
Western Division
La Jolla, California

ABSTRACT

The Navy's Tactical Aircrew Combat Training System (TACTS) provides the sophisticated instrumentation required to obtain in-flight measures of air combat maneuvering (ACM) performance. Empirically based performance assessment methods have been derived using measures from TACTS. Results of this research have been used to develop a measurement framework which may be appropriately applied to estimate overall air combat training effectiveness as well as to provide diagnostic performance analysis of air combat tasks. Clearly, the availability of such objective performance data provides unique opportunities for research validation in the aviation psychology field. In addition, performance assessment methods thus far developed have been incorporated in an automated measurement system called the Performance Assessment and Appraisal System (PAAS). PAAS provides performance based training feedback to operational aircrews in the form of computer generated graphics.

INTRODUCTION

Background

The selection and training of military aviators, and problems associated with the acquisition and retention of flying skill have occupied aviation psychologists for over 30 years. A major difficulty encountered in this line of research has been the lack of objective performance criteria against which to validate ongoing research in aviation selection and training. Obtaining satisfactory criterion measures emerged as a significant research problem early in aviation psychology's history. For example, the principal investigator of the World War II Army/Air Force aviation selection program concluded in his study that "obtaining satisfactory criterion measures was the most fundamental and difficult problem in the Army's aviation psychology program" (Thorndike, 1947, p. 29).

In a more recent review of the aviation criterion measurement problem, North and Griffin (1977) conclude that obtaining valid and reliable criterion measures is still a primary problem in aviation psychology today. In this same review, North and Griffin also state that advancements in technology, such as the Navy's Tactical Aircrew Combat Training System (TACTS) "have the potential to identify and reliably measure relevant . . . human attributes which may provide more accurate and valid predictors of aviator operational performance" (p. 50).

Traditionally, the use of instructor ratings has provided the only means to evaluate the utility of aviator selection methods and to estimate aircrew training progress and proficiency. The availability of in-flight data from instrumented ranges like TACTS offers the possibility of replacing subjective evaluation techniques, commonly used in aviation training, with an objective in-flight measurement approach (Ciavarelli & Brictson, 1978).

TACTS

The recent growth of computing and instrumentation systems, exemplified by the Navy's Tactical Aircrew Combat Training System (TACTS), and various instrumented flight simulators, has permitted the use of in-flight performance measurement systems. The TACTS, formerly known as the Air Combat Maneuvering Range (ACMR), is a sophisticated training facility now in use to train fighter aircrews in air-to-air combat. The system is presently being expanded to train other combat missions as well, e.g., air-to-ground bombing and electronic warfare. TACTS provides data display features which greatly enhance air combat debriefs, and provide a rich source of continuously recorded quantitative data. Design features of TACTS include real-time tracking of aircraft, video tape playback of flight history data, and computer generated weapon launch outcomes (NASC, 1975).

The principal advantage of TACTS is that it uses operational aircraft in a specified airspace which fly highly realistic mock dog fights, while results of these air combat engagements are monitored and recorded. TACTS provides researchers with a rare opportunity to obtain objective "on-the-job" measures of air combat performance. Working under contract to the Naval Aerospace Medical Research Laboratory and the Naval Training Equipment Center, we have seized this opportunity to develop and apply operational measures of air combat performance (Ciavarelli & Williams, 1980).

Measurement Research

Although advanced instrumentation on TACTS has provided the potential for enhanced debrief through its playback features, the system was not designed to accumulate, process, and retrieve information required to determine statistical performance trends which are required to develop Air Combat Maneuvering (ACM) performance criteria. Furthermore, TACTS data outputs are by and large based upon engineering considerations for providing aircraft tracking, cockpit instrument, and weapon launch data. These particular TACTS outputs are essential to monitoring live air combat and for playback of ACM engagements flown, but do not in any sense represent a systematic measurement framework (Ciavarelli, Pettigrew, & Brictson, 1980).

TACTS outputs are essentially raw data, and represent (at best) potential candidate measures for application in the development of TACTS performance criteria. For TACTS to realize its full potential as an aircrew training and performance assessment system, statistically valid and reliable performance criteria must be developed. The next section of this paper outlines a technical approach to the development of air combat performance criteria.

METHOD

Previous Research

Detailed descriptions of research methodology have been presented in a series of technical reports, and can only briefly be summarized here. In our first report (Brictson, Ciavarelli, & Jones, 1977) an analytic framework for conducting criterion measurement research was discussed. This analytic framework described the TACTS as a system which includes *aircrews*, *instructors*, *aircraft*, *weapons*, *mission* and *operating environment*. Our technical approach calls for obtaining performance measures within a system framework which identifies the above TACTS elements, defines the training mission and the operational environment, and provides measurement methods sensitive to the influence of variance identified in the total TACTS system. Each set of performance data collected on TACTS is described and referenced using this systems framework (i.e., training mission, system elements and operating environment), so that performance variations can be related back to specific elements and operating conditions. If desirable, the combination of elements (aircraft, weapons, aircrews, etc.) and the training mission can be systematically controlled to assess their influence on system performance. Most recently, we have developed computer software architecture based on our initial systematic framework for organizing performance data under relevant system variables. Using this computerized data base, a researcher can select and test subsets of performance data by specifying *aircrew*, *aircraft*, *weapon*, *mission* and other specific variables.

A second technical report (Brictson, Ciavarelli, Pettigrew, & Young, 1978) described measurement criteria and assessment methods for evaluating Navy aircrew missile envelope recognition performance. Two generic measurement methods were used to assess aircrew performance in missile envelope recognition. First, a criterion referenced assessment method was applied to score missile launch success (% hits) and to compute task accuracy measures (error from prescribed missile launch boundary). A pilot must fire his weapon within this prescribed boundary to obtain a hit, and the distance from a proper launch window can be used to compute task accuracy (error scored) measures. Second, norm referenced measures, based on empirical data (mean ± standard deviation), were used to evaluate performance related to group standards.

A combination of norm referenced and criterion referenced measures have been proposed for other specific air combat tasks, i.e., radar search and acquisition, visual search and acquisition, tactics and maneuvers, etc. (Ciavarelli, Pettigrew, & Brictson, 1980).

Another technical report (Ciavarelli, Brictson, & Young, 1979) represents a synthesis of the above research methods, and also presents newly introduced methods for evaluating performance results on other critical air combat tasks. Measurement methods for assessing performance on search and locate tasks (initial radar and visual contact with an adversary) and methods for scoring overall engagement outcomes are discussed. A preliminary measurement system has evolved and is based on an integrated set of performance measures related to overall air combat success (win, loss and draw statistics). Methods developed thus far have concentrated on the use of measures related to discrete air combat tasks such as radar and visual acquisition, weapon envelope recognition and engagement outcomes. Additions to this initial measurement schema are planned for the immediate future and include measures related to air combat maneuvering tasks. For example, air combat maneuvering (ACM) engagement state models (Simpson & Oberle, 1977) and algorithms for measuring energy maneuverability performance (Moroney et al, 1979) represent two possible additions in the final formation of an overall ACM measurement system.*

Measurement Framework

Figure 1 shows a simplified air combat maneuvering (ACM) sequence. ACM mission phases are written at the top of the figure, while critical points of measurement are indicated in boxes. Solid boxes represent points for which we now have measures available, and dotted boxes are those measurement points which we will be adding during later phases of the measurement program. Each of the boxed measurement points further represents a specific air combat training objective. Table 1 gives the short form definitions of these training objectives and also indicates their corresponding candidate performance measures. More detailed definitions for training objectives and candidate measures are reported elsewhere (Ciavarelli et al, 1980).

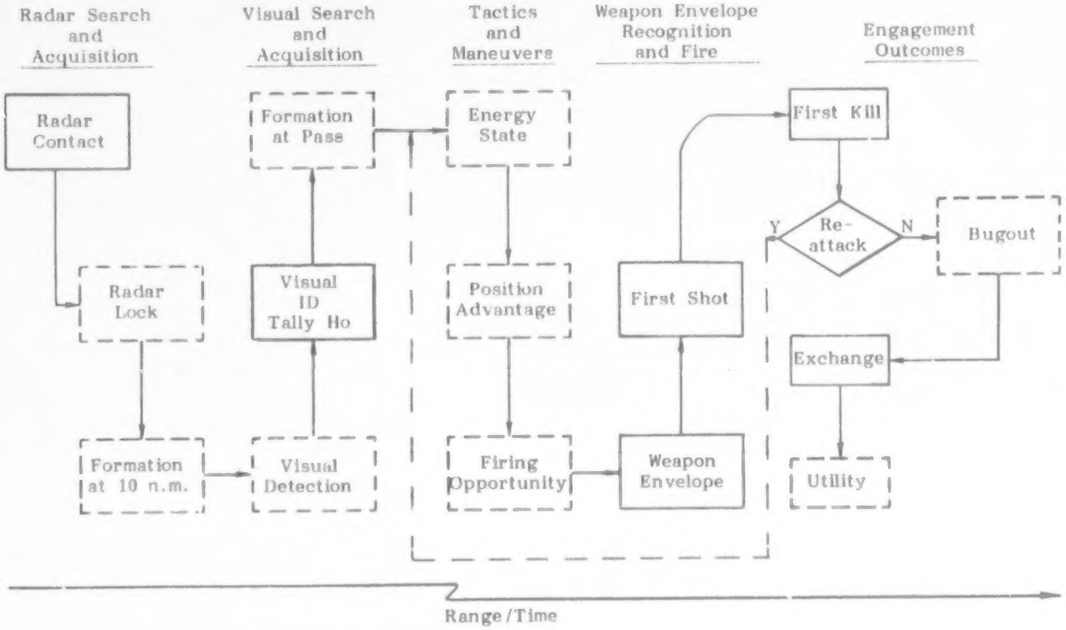

Radar Search and Acquisition
Visual Search and Acquisition
Tactics and Maneuvers
Weapon Envelope Recognition and Fire
Engagement Outcomes
Radar Contact
Radar Lock
Formation at 10 n.m.
Visual Detection
Visual ID Tally Ho
Formation at Pass
Energy State
Position Advantage
Firing Opportunity
Weapon Envelope
First Shot
First Kill
Re-attack
Y
N
Bugout
Exchange
Utility
Range/Time

Figure 1. Simplified air combat engagement sequence: shows engagement phases and measurement points

---

*Measures for engagement state models are based on obtaining a position advantage, i.e., by measuring an aircraft's proximity to the lethal zone of an adversary aircraft. Measures for energy maneuverability are related to optimizing airspeed for efficient air combat maneuvering, i.e., flying aircraft to its aerodynamic ideal.

TABLE 1

Air Combat Engagement Analysis

Training Objective	Performance Measure
1. Obtain early RADAR CONTACT and lock-on	Interaircraft RANGE at radar contact
2. Determine adversary ATTACK FORMATION at 10 n.m.	QUANTITY and POSITION of enemy aircraft
3. Obtain early VISUAL DETECTION of adversary aircraft	Interaircraft RANGE and SUCCESS RATE (%) over engagements flown
4. Obtain early VISUAL IDENTIFICATION of adversary aircraft	Interaircraft RANGE and SUCCESS RATE (%) over engagements flown
5. Determine ATTACK FORMATION at initial pass	QUANTITY and POSITION of enemy aircraft
6. Maintain HIGH ENERGY state[a]	INDICATED AIR SPEED and ALTITUDE (energy package)
7. Gain/maintain POSITION ADVANTAGE[b]	% or proportion of engagement in OFFENSIVE, DEFENSIVE states
8. Gain FIRING OPPORTUNITY[b]	TIME and/or % in envelope or FATAL OFFENSIVE state
9. Obtain FIRST SHOT of engagement	Elapsed TIME and % first shots
10. Obtain FIRST KILL of engagement	Elapsed TIME and % first kills
11. Execute SUCCESSFUL RE-ATTACK	Iterate 6-10 above
12. Execute SUCCESSFUL BUGOUT by staying neutral, maintaining energy, and completing disengagement with no friendly loss	% NEUTRAL, INDICATED AIRSPEED and ALTITUDE, % LOSS at bugout
13. Obtain favorable EXCHANGE (EXCH) RATE	RATIO of fighter to adversary kills
14. Satisfy mission (UTILITY) requirements	NEUTRALIZE threat aircraft and SURVIVE or MINIMIZE losses

[a] See Moroney, Pruit, & Lau, 1979.

[b] Simpson & Oberle, 1977.

Measurement Validation

Some preliminary validation tests of our measurement approach have been completed and are described in the Results section of this paper. More extensive validation procedures, using multivariate statistical methods, are planned in order to understand more fully the composition of critical ACM tasks, their interdependence, and relationship to successful air combat.

RESULTS

Research Summary

Research results from the aforementioned studies are far too extensive to report here in any detail. We have, however, selected some exemplary findings to illustrate some key concepts used in our measurement development (Ciavarelli & Williams, 1980). A radar controlled (vectored) air combat engagement, such as those flown on TACTS, can be conveniently analyzed into a best and worst case situation. A best case situation, according to Navy tactical doctrine, calls for a fighter aircrew to obtain an early radar contact, followed by an early visual sighting (Tally Ho). Once the adversary is sighted, the fighter pilot maneuvers his aircraft to obtain a first shot opportunity and a quick "kill." Conversely, in the worst case situation both radar, tally, and first shot opportunity are lost--and most likely the fighter (friendly) aircraft is "killed." Figure 2 illustrates the best and worst case situations for an air combat engagement. The best and worst case comparisons are somewhat analogous to a gunfight, or duel, where first sight and first shot improves chances of a win. We have tested these doctrinal, or theoretical principles, using TACTS data from about 200 air combat engagements. The results of χ^2 contingency tests demonstrated highly significant relationships ($p<.001$) between radar contact and likelihood of Tally Ho, and between first shot opportunity and likelihood for winning an engagement.

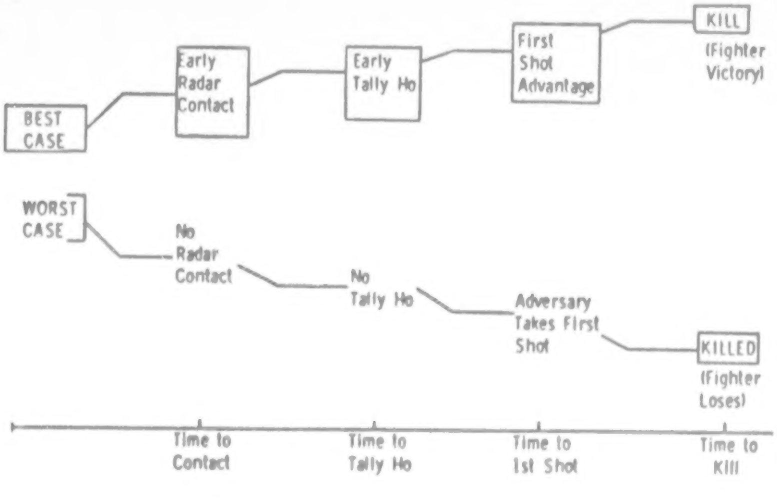

BEST CASE
Early Radar Contact
Early Tally Ho
First Shot Advantage
KILL
(Fighter Victory)
WORST CASE
No Radar Contact
No Tally Ho
Adversary Takes First Shot
KILLED
(Fighter Loses)
Time to Contact
Time to Tally Ho
Time to 1st Shot
Time to Kill

Figure 2. The air combat model: best and worst cases

Figure 3 shows the relationship between radar contact and Tally Ho, while Figure 4 illustrates the relationship between first shot opportunity and win/loss outcomes. Results from this analysis, and other more extensive statistical tests, provide evidence of an agreeable match between theory (ACM tactical doctrine) and practice (performance results). Statistical tests, relating theory and practice, represent a small part of the long-range effort required to demonstrate the validity of our measurement approach.

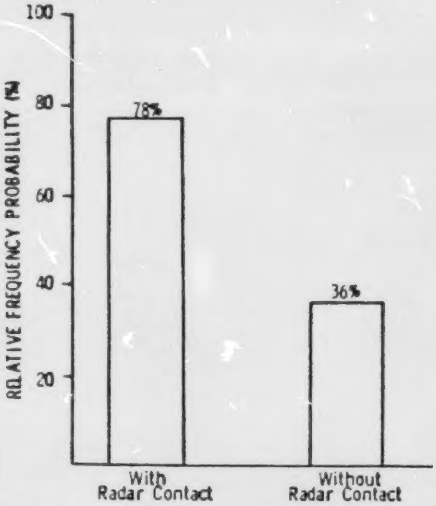


Figure 3. Effect of radar contact on visual acquisition (first Tally Ho)

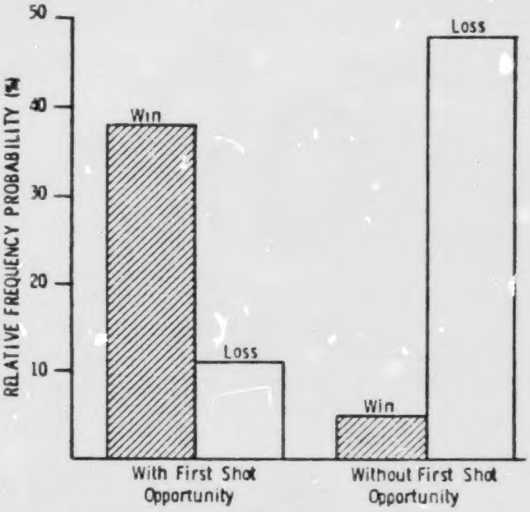


Figure 4. Win and loss rates with and without first shot advantage

A similar research strategy has been used to derive methods for assessing envelope recognition performance. Figure 5 shows the application of a criterion referenced assessment method used to evaluate missile envelope recognition performance between two groups of Navy aircrews.

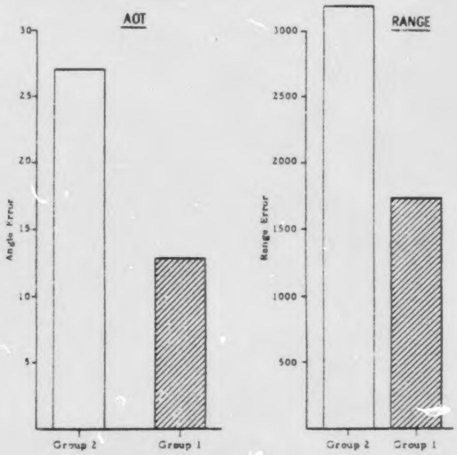


Figure 5. Criterion referenced measures: angle-off-tail (AOT) and range error scores for two groups of F4 pilots

Envelope recognition task accuracy scores depicted in Figure 5 are based on error or deviation from a prescribed boundary (deviation from required range and angle-off-tail), hence these particular measures are criterion referenced.

In Figure 6, an example of a norm referenced measure is provided. Figure 6 presents a polar array of missile fire points plotted in terms of range from target (0-12 Kft) and target offset (0°±90°). An empirical envelope, based on mean ($\overline{X}$) ± one standard deviation (σ) is superimposed on the figure, and used to compare performance of these two groups of aircrews. The empirical envelope represents the application of a norm referenced measurement approach. Both criterion referenced and norm referenced measures are used by operational aircrews to estimate performance related to effective use of weapons. Clearly, an accurately delivered weapon is more likely to find its target. Using both criterion and norm referenced measures, aircrews are able to evaluate their shooting accuracy respectively against doctrine and group standards. Research, previously cited, has demonstrated that these measurement approaches are useful in discriminating performance differences among aircrew members of varying skill levels in weapon delivery accuracy.

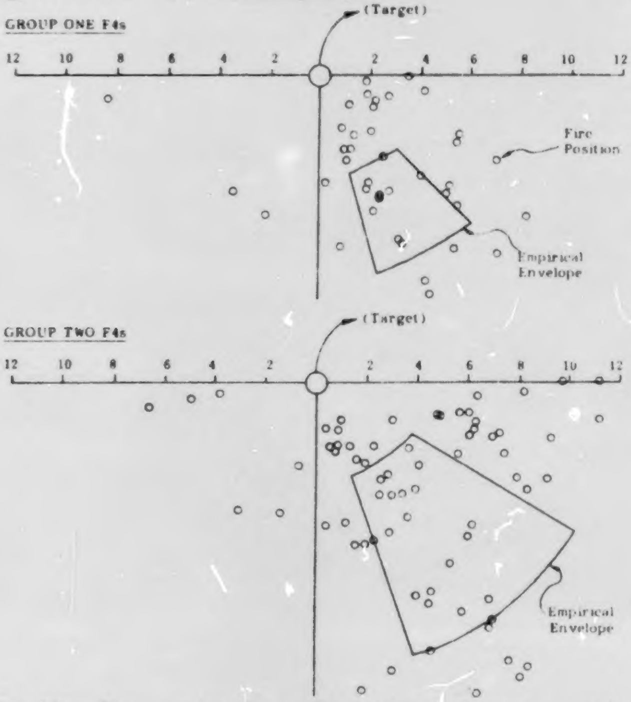


Figure 6. Missile fire points as function of range and angle-off-tail: with norm referenced empirical envelope ($\overline{X}$ $\pm 1\sigma$) superimposed for two group comparison

Our ongoing research in air combat maneuvering measurement has thus far been successful in establishing a systematic framework for analysis, and in selecting candidate ACM measures that make good sense on theoretical and practical grounds. Although much work remains to complete a total ACM measurement system, and to further validate the measurement process, we have already begun to apply our measurement methods directly to operational training.

Operational Application

The main thrust of the air combat measurement research program has been, and continues to be, the development of valid and reliable operational performance criteria. A primary application of such performance criteria is their use in validating ongoing research related to aviator selection, training, and biomedical studies. As the measurement program progressed, however, it became apparent that the performance measurement methods under development would be of immediate benefit to the operational aircrews using the TACTS. Following an extensive analysis of current performance of Navy aircrews using TACTS data, it was determined that substantial improvement to training could be attained only by applying state-of-the-art training and performance assessment methods (Ciavarelli et al, 1980). A computer based air combat debrief system was designed and a baseline system has already been developed. The resulting debrief system, referred to as the Performance Assessment and Appraisal System (PAAS), incorporates performance criteria and assessment methods developed during our criterion research program. The debrief system is designed to accumulate, store, retrieve, and display statistically summarized performance results from TACTS training sessions. Data are provided to aircrews, in an understandable format, for review and diagnostic performance evaluation of air combat training progress. A description of the first design module of the PAAS is presented below.

Performance Assessment and Appraisal System (PAAS)

Overview. On-site observations made during the TACTS measurement research program and verified by data collected from over 300 engagements reveal that TACTS training missions and their associated debriefs vary considerably in training emphasis, content, and quality. Additionally, cumulative data is not routinely collected, so that performance review of more than one mission at a time is not readily available. Determining ACM training progress on TACTS is therefore a time consuming and expensive process in which data from many missions must be hand-compiled. On the basis of our observations and research, and in response to requests by ACM training personnel, we proposed the development of a computerized system called the Performance Assessment and Appraisal System (PAAS), to aid in immediate aircrew debrief and to facilitate longer term evaluation of training progress (Ciavarelli et al, 1980).

The program to design and develop PAAS was sponsored by the Naval Training Equipment Center on the basis that the PAAS would be used as a training debrief aid in conjunction with TACTS. Research, reviewed previously, demonstrated that the air combat mission is a highly structured event, and overall ACM success depends on success at several critical points. With the identification of these phases and empirical results from ongoing measurement studies, we were able to move from the research arena into applications, and the computer based aircrew debrief program was designed. The purpose of PAAS is two-fold. First, it aids in structuring and standardizing the aircrew debrief by providing feedback for only the critical ACM phases identified by our earlier research (see Figure 1). And second, training progress can be estimated because the system stores data from past missions and builds a cumulative data base. We have developed the first module to be included in the PAAS system (Ciavarelli & Williams, 1980).

The currently completed module, Performance Assessment and Combat Effectiveness (PACE), is designed to provide performance review at the squadron level. Recent improvements to PACE graphic formats have been made and these revised computer graphic formats are presented here. In addition to PACE, we are currently developing three more modules to complete the proposed PAAS system: one module to review individual pilot performance (TAD: Training Appraisal and Diagnostics), another to review fleet normative data (NORM: Numerical Operational Readiness Measurement), and a third to provide a complete set of operating procedures (SOP: Standard Operating Procedures). Figure 7 shows the conceptual design configuration of PAAS, with the *now completed* PACE module highlighted. Since only the PACE module of the debrief system is complete, we will describe this particular module in more detail.

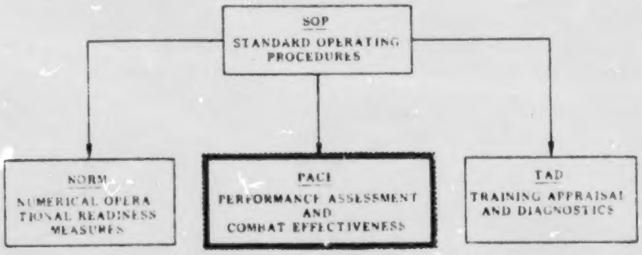

SOP
STANDARD OPERATING PROCEDURES
NORM
NUMERICAL OPERATIONAL READINESS MEASURES
PACE
PERFORMANCE ASSESSMENT AND COMBAT EFFECTIVENESS
TAD
TRAINING APPRAISAL AND DIAGNOSTICS

Figure 7. Conceptual design of Performance Assessment and Appraisal System (PAAS)

Performance Assessment and Combat Effectiveness. PACE has been designed primarily for use by the squadron training officer to review performance of a squadron, as a whole, for key ACM tasks. Missions to be reviewed are selected by entering squadron, mission type, adversary aircraft type, and two inclusive dates in response to prompts by the computer. Performance can be reviewed either on a daily basis, or over a more extended period of time. Thus, PACE could be used to review daily training missions flown by the members of a squadron to look for *improvements* in training on a day-to-day basis, or it could be used to *compare* performance among several squadrons. Performance of squadrons for entire, week-long detachments can also be reviewed and compared. Computer graphics feedback is provided by PACE only for the selected ACM phases identified in Figure 1. Figures 8-13 show some example graphic formats for PACE. Computer generated graphics for other PAAS modules, i.e., NORM and TAD, are currently under development. Graphics for these particular modules will permit users of PAAS to make both *longitudinal* and *crossectional* performance comparisons. In other words, comparisons over time for estimating individual training progress and comparisons among fleet units will be possible using the completed PAAS. A brief description of the PACE module graphics is presented below. Similar graphics will be developed for NORM and TAD modules.

Radar Search and Acquisition. Navy tactical doctrine specifies a requirement to quickly locate an adversary through effective use of on-board radar equipment. An early radar contact provides aircrews with essential position data required to facilitate *visual* acquisition of an adversary. Computer graphic information for radar search and acquisition is provided at two measurement points, radar contact and formation at 10 miles. Measurement points were selected on the basis of both analytical and empirical grounds (see Figure 1 for measurement framework). Figure 8 shows performance data format for radar contact range. Note that a criterion cut-off (see dotted line on Figure 8 at 30 nautical miles) can be used to determine performance level against a doctrine standard. Figure 9 represents graphic data which shows whether or not aircrews have been able to establish the adversary attack formation, i.e., whether they can see one, two, or no adversaries on their radar scope.

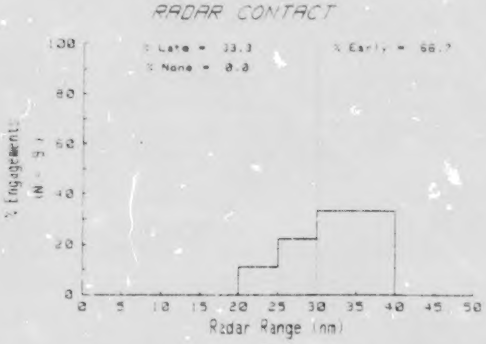


Figure 8. Radar contact, shows percent of engagements for each radar range interval (nautical miles)

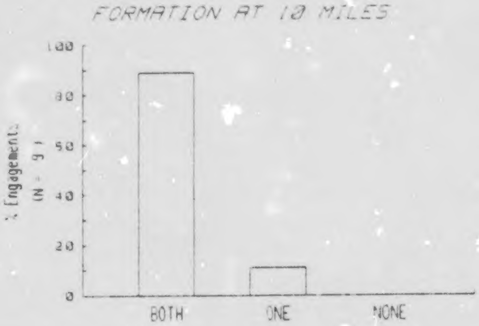


Figure 9. Formation at 10 miles, shows percent of engagements in which both, one, or none adversaries were observed on radar

Visual Search and Acquisition. An important element of air combat tactical doctrine specifies obtaining an initial visual acquisition as early in the ACM engagement as possible. Early detection of your adversary is essential for determining a plan of attack, i.e., selecting a weapon and maneuver to reach opponent's lethal zone. While early identification is essential because most battle scenarios call for positive adversary identification prior to weapon fire. Figure 10 illustrates a typical graphic format used for performance assessment in visual search and acquisition tasks. Graphics for visual detection and close-in attack formation are also available on PAAS, but are not shown here because of space requirements.

Tactics and Maneuvers. In order to get within an adversary's lethal zone, a pilot must first maneuver his aircraft to arrive at the proper weapon fire position. Needless to say, an opponent in air combat has the same goal. The results of this maneuvering between ACM opponents can be analyzed in terms of offensive (+), defensive (-), and neutral (o) engagement states defined in terms of wingline geometry and proximity to an opponent's lethal zone (Simpson & Oberle, 1977). Also, terminal events like the first shot of the engagement and the first kill of the engagement can be used to measure performance related to tactics and maneuvering tasks. Figure 11 illustrates the application of terminal measures by showing percent first shots, percent first kills and their corresponding times. Provisions for adding maneuvering state measures have been made in the design of the PAAS data base.

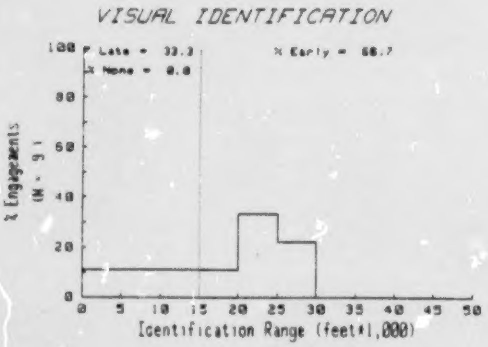


Figure 10. Visual identification, shows percent engagements for each visual identification range (1,000 ft. intervals)

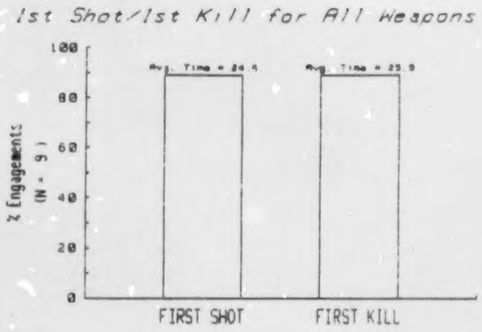


Figure 11. First shot, first kill, shows percent engagements and average time for first shot and first kill

Envelope Recognition. In order to defeat an opponent during an air combat engagement, a pilot must fire his weapon within the boundaries of a launch envelope. As a rule, weapon launch boundaries are specified in terms of interaircraft geometry, such as range and angular limits. Weapon fire success rates and weapon launch accuracy can be determined in terms of interaircraft position data available on TACTS. Figure 12 is an example of one graphic format used for reviewing envelope recognition performance. Success rates for three air intercept missiles (AIM's) are shown. Finally, Figure 13 shows an example of the polar plot format used for determining missile fire position accuracy. Fire position data taken from TACTS are plotted according to whether shots are fired on cockpit side (hot) of target or on belly side (cold) of target. Also, shots can be scored and coded as hits and misses (see Figure 13 for hypothetical data).

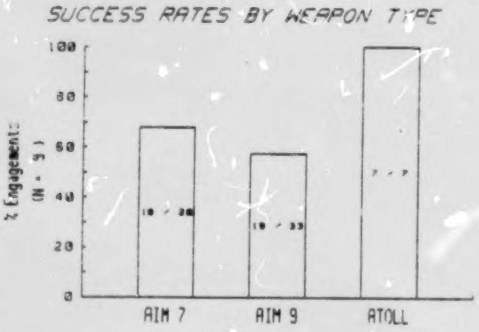


Figure 12. Success rates, shows success rate (kills/total shots) for three missile types

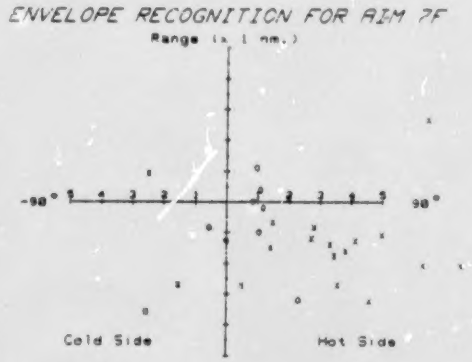


Figure 13. Envelope recognition, shows a sample polar plot of hypothetical AIM 7F fire positions (range in nautical miles, by angles in degrees)

CONCLUSIONS

By way of summary, one system module of the Performance Assessment and Appraisal System has been completed. This particular module, PACE, has been developed for use by the squadron training officer to review performance of a squadron at several mission phases including: radar search and acquisition, visual search and acquisition, tactics and maneuvers, and weapon envelope recognition. Some of the computer graphics designed for use by operational aircrews were presented, and others are under development. Three more modules, TAD, NORM, and SOP will be interfaced with PACE to complete the PAAS program. PAAS will ultimately assist in performance review and feedback for the individual pilot, for squadrons, and for the entire fleet. Although further development of PAAS is underway, the PACE system module in its present form is complete and ready for operational use.

The development and application of operational performance criteria and assessment methods is an essential requirement for effective TACTS training, and should be pursued in other training systems in use. Without a valid and reliable means to assess performance there can be no guarantee that training is in fact occurring. The availability of useful operational measures, when coupled with developments in microprocessing, represents an emerging technology (i.e., automated performance measurement) that promises to greatly enhance the utilization of training systems, like TACTS, through the appropriate application of performance based feedback.

In addition to the direct applications of TACTS performance measures, their potential use to validate ongoing research in aviation psychology has yet to be realized. We are presently undertaking necessary research steps to test the validity and reliability of our measures, so that the research community can view our growing data base as a credible source of objective aviation criterion measures.

NOTES:

The research reported in this paper was completed under Navy contract N61339-77-C-0167 through the Naval Aerospace Medical Research Laboratory, and N61339-78-C-0136 through the Naval Training Equipment Center. The opinions expressed here are those of the authors and do not represent official Department of the Navy policy.

REFERENCES

ACMR Evaluation Summary: AN/USQ-T2(V). Naval Air Systems Command, Report No. ACMR-I-1-75-1, August 1975.

Brictson, C.A., Ciavarelli, A.P., & Jones, T.N. *Development of aircrew performance measures for the air combat maneuvering range (ACMR) (U)* (Report No. L53001). Pensacola, Florida: Naval Aerospace Medical Research Laboratory, June 1977. (CONFIDENTIAL)

Brictson, C.A., Ciavarelli, A.P., Pettigrew, K.W., & Young, P.A. *Performance assessment methods and criteria for the air combat maneuvering range (ACMR): Missile envelope recognition (U)* (Special Report No. 78-4). Pensacola, Florida: Naval Aerospace Medical Research Laboratory, July 1978. (CONFIDENTIAL)

Ciavarelli, A.P., & Brictson, C.A. Air combat maneuvering range (ACMR): Has operational performance measurement entered a golden age? In E. J. Baise & J. M. Miller (Eds.), *Proceedings of the Human Factors Society*, 22nd Annual Meeting, Detroit, October 1978.

Ciavarelli, A.P., Brictson, C.A., & Young, P.A. *Development and application of performance criteria and aircrew assessment methods for the air combat maneuvering range (ACMR) (U)* (Special Report 79-5). Pensacola, Florida: Naval Aerospace Medical Research Laboratory, September 1979. (CONFIDENTIAL)

Ciavarelli, A.P., Pettigrew, K.W., & Brictson, C.A. *Development of a computer based air combat maneuvering range debrief system: Interim report* (Vol. I) (Report No. 78-C-0136-1). Orlando, Florida: Naval Training Equipment Center, January 1980, in press.

Ciavarelli, A.P., & Williams, A.M. *Application of performance based feedback to air combat training*. Paper presented at the 1980 Aerospace Meeting of the Society of Automotive Engineers (SAE Technical Paper Series 801182), Los Angeles, October 1980.

Moroney, W. F., Pruit, R., & Lau, C. *Utilization of energy maneuverability data in improving in-flight performance and performance in air combat maneuvering*. Paper presented at the meeting of the Human Factors Society, Boston, November 1979.

North R. A., & Griffin, G. R. *Aviator selection 1919-1977* (Special Report 77-2). Pensacola, Florida: Naval Aerospace Medical Research Laboratory, October 1977.

Simpson, W. R., & Oberle, R. A. *The numerical analysis of air combat engagements dominated by maneuvering performance* (TM 77-2SA). Patuxent River, Maryland: Naval Air Test Center, June 1977.

Thorndike, R.C. *Research problems and techniques* (Report No. 3). Army Air Forces Aviation Psychology Program Research Reports, Washington, D.C.: U.S. Government Printing Office, 1947.

MEASURES OF EFFECTIVENESS IN EVALUATING A PROTOTYPE GENERAL AVIATION IN-FLIGHT SIMULATOR

Barry Strauch
Embry-Riddle Aeronautical University
Daytona Beach, Florida 32014

Abstract

Traditional measures of evaluating simulator effectiveness are reviewed in relation to the unique needs of general aviation flight simulation. Two measures, simulator-related motivation and safety effectiveness are suggested as methods of examining general aviation simulator effectiveness. Simulator-related motivation refers to the type and extent of motivation pilots have to use simulators in lieu of alternative methods of training, while safety effectiveness refers to the impact of simulator use on the safety of all airspace users. These measures are related to the evaluation of a prototype general aviation simulator, one that combines aspects of a desk-top flight trainer within a popular light general aviation aircraft.

MEASURES OF EFFECTIVENESS IN EVALUATING A PROTOTYPE GENERAL AVIATION IN-FLIGHT SIMULATOR

Flight simulation today is used extensively in all types of aviation training, to assist pilots to fly aircraft as diverse as light general aviation trainers to complex military and commercial aircraft. This popularity is not difficult to explain since simulators:

> ....permit close observation of pilot performance and immediate feedback which improves learning; they permit training pilots in many types of malfunctions not often encountered in flight; they are safe and permit training independent of weather, air traffic and the availability of aircraft; they save fuel, ammunition, targets,(in military aviation) wear and tear on airplanes and, above all, the lives of pilots (Orlansky and String, 1979, p. 9)

Technology has allowed simulator fidelity to advance to the point where pilots need little or no in-flight training to qualify for command of high performance aircraft by reproducing in real time, the motion, visual and kinesthetic cues encountered in actual flight (cf. Huff & Nagel, 1975) As Orlansky and String note, although quite expensive, very sophisticated simulators can be amortized in as little as nine months, depending on the extent of use, by the resultant savings in reduced in-flight training time.

Methods of examining simulator effectiveness have advanced as have the simulators themselves. Provenmire and Roscoe (1971) for example, proposed transfer effectiveness (TE), based on basic learning principles, to assess simulator usefulness as a trainer. A transfer effectiveness ratio (TER) was then outlined to quantify the amount of skill transfer from simulator to aircraft. This can indicate the amount of aircraft time needed to reach a minimum performance level, following simulator training. If simulator-trained pilots require one less hour of flight time compared to aircraft trained pilots for every hour they spend in the simulator, the TER would be one. Orlansky and String (1979) calculated the TER's of simulator-trained pilots from data reported in 22 studies carried out from 1967 to 1977 and obtained ratios from -0.4 to 1.9, with a median of 0.48. This indicates that simulator trained pilots can expect to save about half an hour of in-flight training for every hour they spend in the simulator. Cost-effectiveness of the simulators, the extent to which their use results in cost savings, can then be calculated, once the costs of simulator and aircraft training time are known.

In a later study, Provenmire and Roscoe (1973) examined the incremental transfer effectiveness ratio (ITER) of flight simulators, based on an inverse relationship known to exist between the amount of prior learning and the extent of transfer that will thereafter take place. The amount of flight time saved by simulator training can be expected to decrease with increasing simulator training, as the learner will approach asymptote on learning simulator-based flight skills.

Subjective methods have also been used to evaluate simulator effectiveness, often in conjunction with more objective measures of transfer and cost effectiveness. These generally employ experienced pilots to judge the qualities of flight simulators according to predefined measures, generally concerning the realism and fidelity of simulators. As with many rating scales, subjectivity of responses can be controlled by the type of scale employed. Other design considerations help to insure scales of high psychometric quality.

395

Regardless of the methodology, research demonstrating simulator effectiveness has been remarkably consistent over the years. As Orlansky and String (1979) found, simulators are repeatedly shown to be cost effective substitutes for in-flight pilot training. In fact, recent research demonstrated that only partial fidelity to actual aircraft environments is needed for simulators to be shown effective. Trollip (1979) showed that low time general aviation pilots can learn complex instrument maneuvers by a computer graphic display of an aircraft instrument panel and an associated computer assisted instruction course. Lintern (1980) found that after adding minimal visual cues to a Singer-Link GAT-2 general aviation simulator, flight naive pilots can safely land a single engine aircraft.

So consistent is the literature on the training effectiveness of flight simulators that some researchers have suggested that future simulator research be directed to other, more productive areas. Caro (1973), for example, argues that simulator research and development should stress as much the "organization and content of the (pilot) training programs" as well as the hardware and software of simulation systems. Adams (1979) goes so far as to suggest that evaluation of the training effectiveness of simulators is unnecessary as long as the particular simulator incorporates the principles of learning contained in the literature. He states that:

> .... a system built on sound scientific laws needs less concern with evaluation because a good scientific law produces accurate prediction, and when the outcome can be predicted it is redundant to conduct an evaluation (p. 717).

Adams and Caro both rely on extensive research and development in commercial simulation in developing their arguments; however, in general aviation simulator research and development has lagged behind those areas. This is primarily because general aviation simulators have been used sparingly as both trainers and research vehicles in comparison with those used in military and commercial aviation. The demand for general aviation simulation training has not warranted the substantial expenditures of effort and expense required to develop sophisticated simulators since the cost of basic in-flight training was relatively inexpensive. In commercial aviation, on the other hand, where a Boeing 747 costs nearly $5,000 an hour to operate, annual savings to an airline from simulator use can amount to several million dollars.

General aviation, however, is different. No single user will purchase a billion dollar fleet of aircraft. Sophisticated simulators, shared by many purchasers, are available only to pilots top-of-the-line turbojet aircraft. For lower ends of the general aviation product line, the cost effectiveness of development and acquisition has not justified large scale simulator research. Although Lintern (1980), Roscoe (1980), and Trollip (1979), for example, have demonstrated that high-fidelity simulators are not needed for effective flight training, in general aviation the simulators available, while economical, are often so unrealistic as to discourage large scale use. Consequently, the average personal use general aviation pilot selecting a method of training has often considered mostly (not insubstantial) economic factors in deciding to use simulators in place of in-flight training. For professional pilots, this has not been true since simulator costs are borne by other or included within other legitimate business expenses. Underlying this distinction is that fact that simulator training is voluntary in general aviation and mandatory in military and commercial aviation. Although Caro, Shelnutt and Spears (1981) showed that even within the military, motivational factors can still influence the effectiveness of simulation training the fact remains that in general aviation the more motivational factors, beyond exclusively economic ones, influencing the pilot to use a simulator in place of an aircraft, the more simulators will be used.

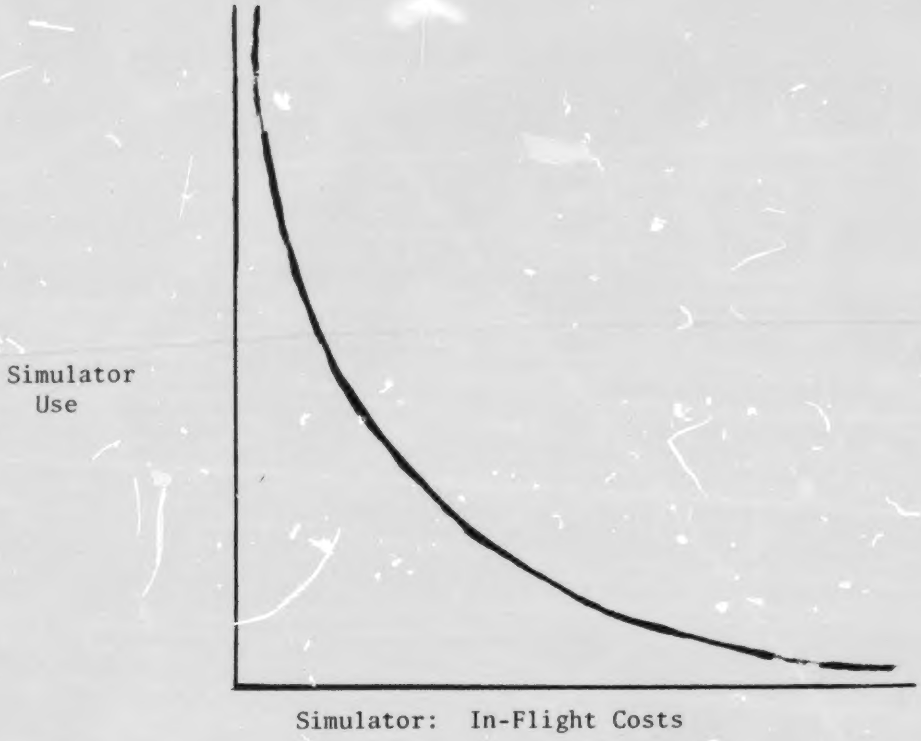
As a result, researchers concerned with simulation effectiveness should develop new measures to examine that effectiveness as it relates to the particular needs and missions of general aviation. This paper outlines two such measures and their application to a prototype general aviation simulator.

SIMULATOR-RELATED MOTIVATION

For general aviation, simulator realism is not as important as what that realism can be expected to bring about, that is, motivation to use the simulator. By extrapolating from findings in military and commercial simulation research, general aviation simulation considerations have overlooked the fact that the simulators themselves must provide motivation to the pilot in order for him or her to use them. While economic factors have been and continue to be major issues in general aviation simulator use, unless the simulators themselves provide motivation to the pilot, he or she will use them primarily because they are less expensive than in-flight training. Therefore, when economic factors are no longer important, the pilot will be less willing to use them.

This can be portrayed by Figure 1, which illustrates the hypothesized demand curve of simulator use as a function of the cost differential between simulator and in-flight training. This differential, expressed as a ratio of simulator: in-flight costs, is inversely related to simulator use, i.e., as the ratio increases simulator use will decrease, all other considerations being equal.

Figure 1



Proposed Demand Curve of Simulator use as a function of simulator: in-flight costs

Intrinsic motivation, a psychological construct long used in social, industrial and educational psychology, has not been applied to the study of general aviation simulation. Yet, while considered a powerful behavioral variable in other fields, its influence in simulation research has been minimal. It is suggested that, all other considerations being equal, the greater the motivation that a simulator can create for its use, the more it will then be utilized.

The mechanism by which this motivation can be engendered in the general aviation pilot is a subject for future research. Simulator realism or fidelity, appears to be a reasonable variable. The more realistic the simulator, the more it can be expected to be used. Other factors, equally powerful, can also be proposed. The extent to which a pilot believes that a simulator will enhance piloting ability in ways that an aircraft cannot and the degree to which simulator use is popular among community pilots, are suggested motivational variables. Simulator-related motivation will be dynamic according to the environment. However, the nature of the simulator-related motivation is not as important a variable in influencing simulator use as the strength of that relationship. Thus, if the simulator:in-flight cost ratio increases beyond a certain point, simulator capabilities, to insure continued simulator use, should enhance the motivation of a pilot use it in place of an aircraft. Research is needed to examine the nature of simulator-related motivation to determine how simulators can be developed, and simulator training curricula designed, to enhance the desire of the general aviation pilot to use a simulator as a trainer in place of in-flight training.

SAFETY EFFECTIVENESS

Second and no less important in the consideration of simulator effectiveness is the issue of safety. Simulators have long been recognized for their ability to safely train pilots in maneuvers which could otherwise jeopardize the safety of an actual flight. In addition, as the recent tragic San Diego mid-air collision illustrates (NTSB, 1979) two aircraft in the same airspace can and sometimes do collide with one another, however remote the probability of such an incident occurring. Simulator use in place of in-flight training can enhance the safety of all airspace users if for no other reason than because it reduces the number of aircraft sharing a given air space.

The response of the Federal government to that incident was to restrict user access to crowded airspace as well as encourage pilots to pursue instrument flight training at satellite airports. The FAA has been actively expanding the facilities at those airports, an expensive proposition at best and one which overlooks the fact that diverted air traffic still poses hazards to other air traffic, albeit in different airspace. The extent to which general aviation flight simulators, by their use, can ease the crowding of any airspace should be an essential consideration in examining simulator effectiveness. It is proposed then that "safety effectiveness" be considered along with motivation and other measures in determining the utility of flight simulators.

Safety effectiveness is, of course, integrally related to the intrinsic motivation the simulator can provide to the pilot. Safety effectiveness is directly influenced by simulator-related motivation since the more motivation the simulator provides the more one can predict that a pilot will use a simulator.

The more the simulator is used in place of in-flight training the less crowded the airspace will be and the smaller the overall likelihood of a collision occurring. The determination of safety effectiveness is an attempt to account for the effect simulator use will have on overall aviation safety.

Safety effectiveness is proposed to be a function of two separate but related factors: 1) the degree to which pilots gain proficiency through increased simulator use and 2) the extent to which airspace crowding is reduced in a given airspace. This can be portrayed by three variables, interacting in a manner uncertain at present:

TE = Transfer Effectiveness.

ΔA = Change in the number of aircraft in a given airspace.

T = Time spent in a simulator.

A hypothetical unit, safety effectiveness, is proposed to account for the safety enhancement accruing to all airspace users by general aviation simulator utilization. The relationship between simulator use and simulator transfer effectiveness has been discussed. The higher the transfer effectiveness, of course, the more one can predict that simulator use will enhance pilot proficiency in an aircraft. At the same time, the more a simulator is used, the less time an aircraft will occupy a given airspace. Since any two aircraft in the same airspace have a probability of colliding with one another, however remote that probability, then the greater the number of aircraft in an airspace, the higher the overall likelihood that any two of those aircraft will collide. Therefore, simulator use, by the fact that it occurs in place of aircraft occupying an airspace, is enhancing the safety of all airspace users by reducing the total number of aircraft in the airspace and the overall probability of an aircraft collision.

Of course, many factors enter into the probability of a ground or mid-air collision, including aircraft speed and maneuverability, type of radar coverage and air traffic control available to and utilized by airspace users, weather and the like. In addition, advanced radar capabilities and collision avoidance systems expected to be installed in the near future should further reduce the probability of aircraft collisions. But, researchers have yet to determine the role of general aviation simulator use in improving aviation safety by its reduction of the number of aircraft in an airspace.

In addition to reducing air traffic, simulator use increases pilot proficiency as a function of its transfer effectiveness ratio. The more a pilot uses a simulator or an aircraft, the more proficient he or she can be expected to become and as a result, the greater the enhancement to air safety. The relationship of proficiency to air safety has been recognized to the extent that it is incorporated, as minimum pilot flight time for legal flight within the Federal Aviation Regulations (14 CFR 61.57). Its relationship to the hypothesized function, safety effectiveness, needs to be determined through empirical research.

This research would yield the hypothetical function, safety effectiveness, through its determination of the reduced probability of aircraft collisions occurring as a function of the reduction of the number aircraft sharing an airspace resulting from the simulator use. The nature of the relationship of pilot proficiency to safety effectiveness is indeterminate at this point; its ability to be quantified, uncertain. Nevertheless, it is expected that this would have a direct relationship to safety effectiveness. Other variables related to general aviation simulator use no doubt can have an effect on safety effectiveness. It is hoped that the concept of safety effectiveness will stimulate researchers to explore the nature of the relationship of general aviation simulator use to the safety of all airspace users, in addition to the pilots using them.

THE IN-FLIGHT SIMULATOR

These measures were proposed in the evaluation of a unique type of general aviation simulator that was developed by the National Aeronautics and Space Administration (NASA). This simulator integrates elements of an inexpensive desk top simulator, an Analog Training Computer (ATC) 610, into a popular flight trainer, a Cessna 172. This afforded the aircraft the ability to simulate instrument approaches and other instrument flight rules (IFR) maneuvers independent of any ground based navaids. The resultant aircraft is only a partial simulator since it does not simulate flight itself only navigational features required for instrument flight as well as other simulator capabilities such as flight data recording.

This ability meets some of the major limitations of most general aviation simulators. It provides the sensations encountered in actual flight, as well as the ability to reproduce instrument indications of various IFR maneuvers. On the other hand, the simulator, while saving time and therefore expense by obviating the necessity of entering and sequencing into airport traffic patterns, relative to other general aviation simulators would still have poor cost-effectiveness, albeit high transfer effectiveness, since it is contained within an aircraft. As a result, it could be expected that further research and development to determine the applicability and marketability of such a prototype simulator would not be undertaken due to the poor expected economic returns. However, other measures of simulator effectiveness could uncover benefits to its use.

It is suggested that new methods and concepts can measure and suggest additional facets of the simulator's effectiveness. The degree to which it enhances motivation could be measured empirically by observing the extent to which and the reasons why pilots would use the simulator in place of alternative methods of flight training. From behavioral observation then, inferences about the motivation generated by the simulator can be carried out, particularly in relationship to other factors such as the simulator: in-flight cost ratio, aircraft availability and the like.

The extent of pilot use of the simulator would then directly influence its safety effectiveness. This can only be measured by observing the extent of pilot utilization of the simulator in comparison with baseline observations of the extent of in-flight training use by similar pilots in similar environments.

The in-flight simulator is only a prototype simulator, its development unlikely. However, it provided the opportunity to offer a new approach to general aviation simulation that has not been considered by simulation researchers. Perhaps more important, the in-flight simulator pointed to limitations in using measures of simulator effectiveness not developed specifically for the particular general aviation context.

The need to develop new measures of simulator effectiveness is heightened by the impact of advanced microprocessor technology on general aviation simulation. Already general aviation simulators are on the market providing pilots with computer generated visual capabilities. With the emergence of advanced flight displays and all digital systems in state of the art aircraft, the application of this technology to other models of the general aviation line can be expected in the near future. General aviation aircraft and simulator manufacturers will surely capitalize on these advances. At the same time, the general aviation market, aircraft utilization and pilot proficiency are changing as a direct result of political and economic forces in operation. The nature of general aviation simulation in the near future is, as a result, uncertain. Researchers must be alert to the need to develop new measures to evaluate simulator effectiveness, measures that can accurately gauge and uncover the often subtle effects of simulation use. The result will be increased sensitivity to simulator potential and, hopefully, more unique applications of simulator technology to general aviation.

REFERENCES

Adams, J.A. On the evaluation of training devices. Human Factors, 1979, 21, 711-720.

Caro, P.W. Aircraft simulators and pilot training. Human Factors, 1973, 15, 502-509.

Caro, P.W., Shelnutt, J.B., and Spears, W.D. Aircrew Training Devices: Utilization (AFHRL-TR-80-35). Brooks AFB, TX: Air Force Systems Command, January, 1981.

Huff, E.M. and Nagel, D.C. Psychological aspects of aeronautical flight simulation. American Psychologist, 1975, 30, 426-439.

Lintern, G. Transfer of landing skill after training with supplementary visual cues. Human Factors, 1980, 22, 81-88.

National Transportation Safety Board. Aircraft accident report - Pacific Southwest Airlines, Inc. B727, N533PS, and Gibbs Flite Center, Inc., Cessna 172, N7711G, San Diego, CA, September 25, 1978. (NTSB AAR 79-5) Washington, DC: Author, April 20, 1979.

Orlansky, J., and String, J. Cost-effectiveness of flight simulators for military training. (NAVTRAEQUIPCEN IH-316). Orlando, FL: Naval Training Equipment Center, November, 1979.

Provenmire, H.K., and Roscoe, S.N. An evaluation of ground-based flight trainers in routine primary flight training. Human Factors, 1971, 13, 109-116.

Provenmire, H.K., and Roscoe, S.N. The incremental transfer effectiveness of a ground-based general aviation trainer. Human Factors, 1973, 15, 534-543.

Roscoe, S.N. Aviation Psychology. Ames, IA: Iowa State University Press, 1980.

Trollip, S.R. The evaluation of a complex computer-based flight procedures trainer. Human Factors, 1979, 21, 47-54.

1981 SYMPOSIUM ON AVIATION PSYCHOLOGY

Attendee List

Robert A. Alkov, Ph.D.
Naval Safety Center
Naval Air Station
Norfolk, VA 23511

William C. Allen
Stanford University
P.O. Box 10281
Stanford, CA 94035

Mr. Karl F. Anderson
Flight Engineers International
905 Sixteenth St., N.W.
Washington, D.C. 20006

Prof. R. J. Anderson
The Ohio State University
155 South Oval Drive
Columbus, OH 43210

Dr. Tom Amerson
The Essex Corp.
333 North Fairfax
Alexandria, VA

Mr. Charles E. Arnold
FAA
2300 E. Devon Ave.
Des Plaines, IL 60018

CPT. DR. Hellfried Aubauer
Austrian Airlines
Fontanastrasse 1
A-1107 Vienna, Austria

J. P. (Jack) Bailey
Transport Canada
P.O. Box 8550, 125 Garry St.
Winnipeg, Manitoba
Canada R3C OP6

James R. Barney
IBM Corp.
304A P70 Federal Systems Division
Owego, NY 13827

Dr. Sheldon Baron
Bolt, Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02238

Mr. James V. Barry
Executive Air Fleet Corporation
118 Billy Diehl Rd.
Teterboro, NJ 07608

Dr. R. P. Bateman
SRL, Inc.
2800 Indian Ripple Road
Dayton, OH 45440

William A. Beckwith
Georgia State University
University Plaza
Atlanta, GA 30303

Mr. Ivan L. Belvea
Lear Siegler, Inc.
4141 Eastern Ave., S.E.
M/S 128
Grand Rapids, MI 49508

Dr. Russell A. Benei
Essex Corporation
333 N. Fairfax St.
Alexandria, VA

Eugene W. Benson
Systems Research Laboratories, Inc.
2800 Indian Ripple Road
Dayton, OH 45440

Mohamed Rashael Berenji
Saudi Arabian Airline
Saudi Arabia

Dr. Dennis B. Beringer
University of Wisconsin at Madison
Department of Industrial Engineering
1513 University Ave.
Madison, WI 53706

Dr. Jerome I. Berlin
Embry-Riddle Aeronauical Univ.
Aviation Research Center PD-AR
Regional Airport
Daytona Beach, FL 32014

Dr. Gayle Berry
The Ohio State University
Industrial & Systems Engineering
1971 Neil Ave.
Columbus, OH 43210

Reid B. Berry
Coordinator of Development
Moody Aviation
P.O. Box 429
Elizabethton, TN 37643

Dr. Vivek D. Bhise
Ford Motor Company
Automotive Safety Center
P.O. Box 2053
Dearborn, MI 48121

David W. Biers
University of Dayton
Dayton, OH 45469

B. Braune
University of Illinois

Berl Brechner
Executive Editor
Flying Magazine

Dr. Fritz Brecke
Veda, Inc.
Friars Office Prk,
Building E - Suite 320
7851 Mission Court
San Diego, CA 92018

Dave Britton - Crew Station Design Engineer
USAF, Aeronautical Systems Div. - WPAFB, OH
ASD/ENECC
WPAFB, OH 45433

Dr. Richard K. Brown
University of Southern California
Institute of Safety and Systems Management
Office of Extension and In-Service Management
Room 104 University Park
Los Angeles, CA 90007

Mr. William R. Brown
U.S. Army Research Institute
P.O. box 476
Ft. Rucker, AL 36362

Mr. Kevin Burns
ASD/ENECH
W-P AFB, OH 45433

Capt. Gerry Butler - Crew Station Design Engineer
USAF Aeronautical Systems Div. W-P AFB, OH
ASD/ENECC
WPAFB, OH 45433

Dr. Keith Butler
Boeing Computer Services-Consulting Division
919 S.W. Grady Way Mail Stop 6K-01
Renton, WA 98055

Mr. Robert Butler, Mgr. Aircrew Training
United Technologies Cor.
Rentschler Airport
400 Main St.
E. Hartford, CT 06108

Capt. R.A. Byrd, Jr.
Manager - Flight Operations - Technical
Delta Air Lines, Inc.
Atlanta Hartsfield Airport
Atlanta, GA 30320

Denise Byrnes
Naval Air Development Center
Code 6022
Warminster, PA 18974

Ms. Gloria L. Calhoun
Bunker Ramo Corporation
4130 Linden Ave., Suite 302
Dayton, OH 45432

Webb Caster
Aviation Simulation Technology, Inc.
Hanscom Field East
Bedford, Mass.

Dean Catoe, Flight Station Design Engineer
Lockheed-Georgia Company
2059 Winding Creek Lane
Marietta, GA 30064

Courtney Chapman
Department of Aviation
The Ohio State University

Sherry Chappell
NASA Ames/Tufts. University
1075 Space Parkway #151
Mountain View, CA 94043

Ed Cheaney
Battelle Columbus Laboratories
625 Ellis Street, Suite 305
Mountain View, CA 94043

Dr. Jerry M. Childs
Seville Research Corp.
Pensacola, Florida

Gerald P. Chubb
AF Aerospace Medical Research Laboratory
Plans and Programs (TSX)
W-P AFB, OH 45433

Tony Ciavarelli
Managing Scientist
Dunlap and Associates, Inc.
920 Kline St.
La Jolla, CA 92037

Don Condra
Bunker Ramo Corp
41300 Linden Suite 302
Dayton, OH 45432

Capt. Bill Connor
ADJ. Professor
Delta Air Lines/ERAU Graduate School
9420 S.W. 102 Ct.
Miami, FL 33176

Dr. William J. Conroy
State University of New York at Buffalo
Richmond Quadrangle - c/o Housing Office
Ellicott Complex
Buffalo, NY 14261

Ed Cope
Crew Station Design Engineer
USAF, Aeronautical Systems Div. - W-P AFB, OH
ASD/ENECC
WPAFB, OH 45433

Mr. Mark S. Crabtree
Systems Research Laboratories, Inc.
2800 Indian Ripple Rd.
Dayton, OH 45440

Mike Cuddy
Flight Manager
United Airlines
P.O. Box 92245, World Way Postal Center
Los Angeles, CA 90009

Dr. Charles Csuri
Computer and Information Sciences
The Ohio State University

Dr. Diane Damos
Univ. of Oregon
Eugene, Oregon 97405

Jim Davis
Sr. Human Factors Engineer
General Dynamics Ft. Worth Div.
3800 Wedgworth Rd. So.
Ft. Worth, TX 76133

Capt. Joe Davies
ALPA
1223 Robin Hood Drive
Elgin, Illinois 60120

Dr. Alan Diehl
Program Scientist, Human Performance
FAA, Office of Aviation Medicine
800 Independence Ave. S.W.
Washington D.C. 20591

Dr. Theodore J. (Ted) Doll
Georgia Tech
EES/SEL
Georgia Institute of Technology
Atlanta, GA 30332

Mr. David Domino
Aircrew Training Coordinator
United Technologies Corporation
Rentschler Airport
400 Main St.
E. Hartford, CT 06108

Mr. Jack Duffy
Research Assistant
Biotechnology Inc.

J. Steven Eckel
Pilot/Human Factors
General Physics Corporation
Suite 240
1010 Woodman Drive
Dayton, OH 45432

Duane Edelman
Flight Standards
Republic Airlines
16721 Biscayne Ave.
Farmington, Minnesota 55024

Bill Edmunds
Air Line Pilots Association
1625 Massachusetts Ave. NW
Washington, DC 20036

Col. Arne Ellermets
Chairman Air Force Aerospace Studies
The Ohio State University
353 Converse Hall
2121 Tuttle Park Place
Columbus, OH 43210

Lt. Micheline Y. Eyraud
Naval Aerospace Medical Research Lab
NAS Pensacola, FL 32506

Delmar M. Fadden
Boeing
14901 N.E. 147th Pl.
Woodinville, WA 98072

Dipl. Psych. Frank Fehler
Federal German Government
15 Linda St.
Pensacola, FL 32506

John M. Flach
Research Associate- Human Performance Center
OSU
Human Performance Center
404-B W. 19th Ave.
Columbus, OH 43210

Bill Flathers
MITRE Corp.
1820 Dolly Madison Blvd.
Mclean, VA 22102

Dr. Clayton Foushee
NASA Ames Research Center
Moffett Field, CA

Frank D. Fowler
Fowler and Associates
11516 E. Colonial Dr.
Orlando, FL 32803

Rudy Frasca
Pres., Frasca International,Inc.
606 S. Neil
Champaign, IL 61820

David E. Frearson
USAF Flight Dynamics Lab
AFWAL/FIGR
Wright-Patterson AFB, OH 45433

David Fulkerson
Hughes Aircraft Company
6087 Horner St.
Los Angeles, CA 90035

Reginald Furlonger
British Embassy
3100 Massachusetts Ave. NW
Washington, DC 20008

Professor Eugene Galanter
Columbia University
324 Schermerhorn Hall
New York, NY 10027

John B. Galipault
President, Aviation Safety Institute
Aviation Safety Institute
Box 304
Worthington, OH 43805

James M. Galway
Executive Air Fleet Corporation
118 Billy Diehl Road
Teterboro, NJ 07608

Janice M. Gavern
Aero Systems Division Human Factors Branch
ASD/Enech
Wright Patterson AFB, OH

Dr. Walter C. Giffin
The Ohio State University
Department of Industrial & Systems Engineering
1971 Neil Ave.
Columbus, OH 43210

Dr. Donald Glower, Dean
College of Engineering
The Ohio State University

Evelyn Gora, M.A.
Technical University of Munich
Department of Psychology
Lothstr. 17
8000 Munich 2
West Germany

Charles Graham, Ph.D.
Midwest Research Institute
425 Volker Blvd.
Kansas City, MO 64110

Dr. Joanne Green
Georgia Tech - Engineering Experiment
Georgia Tech
EES/SEL
Atlanta, GA 30332

Captain L. D. Griffin
Airline Pilots Assoc.
4737 46th Ave. NE
Seattle, Washington 98105

L. Lee Griffin
AFAMRL W-PAFB
W-PAFB, OH 45433

Dr. J. H. Grosslight
Chairman and Professor
Florida State University
Department of Psychology
Tallahassee, FL 32306

Eric V. Gruber
Research Associate
Embry-Riddle Aeronautical University
Regional Airport
Daytona Beach, FL 32014

David Gunning
Aeronautical Systems Division
Human Factors Branch
ASD/ENECH
Wright Patterson AFB, OH

Chris Hale
Systems Research Laboratories
2800 Indian Ripple Rd.
Dayton, OH 45440

Dr. Keith A. Hall
The Ohio State University
121 Ramseyer Hall
29 W. Woodruff Ave.
Columbus, OH 43210

Captain P. W. (Bill) Hanson
Allied Pilots Assoc.
P.O. Box 5524
Arlington, TX 76011

Richard Harner
Air Force ASD/ENECH
Wright Patterson AFB, OH

Dr. Randall L. Harris, Sr.
NASA, Langley Research Center
Mail Stop 152 D
Hampton, VA 23665

Doug Harvey
Manager, Airmen Research
FAA
Technical Center
ACT-340
Atlantic City Airport, NJ 08405

Douglas M. Hazelwood
Supervisor of Eng. Flight
Wallace Div.
Cessna Aircraft Corp.

Dr. Robert T. Hennessy
National Research Council
2101 Consitution Ave. NW JH 819
Washington, DC 20418

Sam Herron
Human Factors Specialist
Bunker Ramo Corp.
4130 Linden Ave., Suite 302
Dayton, OH 45432

John C. Heurtley
Chief Sys. Test & Eval. Div.
DOT/FAA Technical Center
Atlantic City Airport, NJ 08405

Dr. Lloyd Hitchcock
Federal Aviation Administration
FAA Technical Center
Atlantic City, NJ 08405

Larry Hittinger
Department of Psychology
The Ohio State University

Thomas Hitzeman
Crew Station Design Engineer
USAF Aeronautical Systems Division
ASD/ENECC
WPAFB, OH 45433

Mr. Wesley Holder
Mgr. Corporate Aircraft Department
United Technologies Corporation
Rentschler Airport
400 Main St.
E. Hartford, CT 06108

Dr. Chuck Holmes
E-RAU PD-Af
Regional Airport
Daytona Beach, FL 32014

Helene P. Iavecchia
ANALYTICS
2500 Maryland Rd.
Willow Grove, PA 19090

Joyce H. Iavecchia
Naval Air Development Center
Warminster, PA 18974
Code 6022

Dr. Paul D. Isaac, Acting Chairman
Department of Psychology, OSU
403 Arps Hall
1945 N. High St.
Columbus, OH 43210

Larry J. Ivey
USAF ASD/ENECH
Wright Patterson AFB, OH 45459

Dave Jackson
United Airlines
32nd and Quebec
Denver, CO 80207

Dr. Richard Jagacinski
The Ohio State University
Human Performance Center
404-B W. 17th Ave.
Columbus, OH 43210

Sharon Jago
ASRS Project-Battelle Columbus Labs
625 Ellis St., Suite 305
Mountain View, CA 94043

John Jenkins
United Airlines

Dr. Richard S. Jensen, Director
Aviation Psychology Lab
The The Ohio State University

Michael Karim
Traveling Aviation Seminars
Dublin, OH

Greg Kearsley
HUMRO
300 N. Wash. St.
Alexandria, VA 22314

Lester R. Kerfoot, Jr.
Aviation Safety Consultant
System Safety Associates Ltd.
6216 Apache Street
Springfield, VA 22150

Dr. Sergei Kochkin
United Airlines
P.O. Box 66100
Chicago, IL 60666

Lt. Col. Jeff Koonce, PH.D.
U.S. Air Force Academy, Colorado
DFBL
USAF Academy, CO 80840

Frances A. Kniess
AF Aerospace Medical Research Lab
HEA - Visual Displays Branch
Wright Patterson AFB, OH 45433

Dr. Conrad Kraft
Boeing Aerospace Company
Seattle, Washington

Sue Kuramoto
Human Factors Engineer
Aeronautical Systems Division
USAF; Human Factors Branch
ASD/ENECH
WPAFB, OH 45433

Prof. Bibb Latane'
Department of Psychology
The Ohio State University

Mr. Russell S. Lawton
Assist. Vice President
AOPA
7315 Wisconsin Ave.
Washington, DC 20014

James P. Loomis, Manager
Transportation Systems Section
Battelle, Columbus Laboratories
505 King Ave.
Columbus, OH 43201

Peter B. Lovering
Bunker Ramo Corp.
4130 Linden Avenue
Suite 302
Dayton, Ohio 45432

Amir Mane'
University of Illinois
Department of Psychology
Champaign, Ill. 61820

Sue Mangold
Department of Psychology
The Ohio State University

Wayne Martin
USAF Aero Med Lab
AFAMRL/HEA
WPAFB, OH 45433

Captain Thomas A. McClone, Jr.; Flight Manager
United Airlines
JFK International Airport
Jamaica, NY 11430

David E. Miller
Rockwell International Corp.
4300 East Fifth Avenue
Columbus, Ohio 43216

Kirk Moffitt
Systems Research Laboratories, Inc.
2800 Indian Ripple Rd.
Dayton, Oh. 45440

Dr. Melvin Montemerlo
NASA Headquarters
Washington, D.C.

Richard Moss
Flight Dynamics Lab-WPAFB
1162 Geneva Road
Xenia, Ohio 45385

S. Joy Mountford
Honeywell SRC
Systems & Research Center
Aerospace Defense Group

Robert W. Mudge, President
Cockpit Management Resources, Inc.
P.O. Box 969
Center Harbor, NH 03226

John R. Munger
National Field Coordinator For Priority Air Freight
Suite 235
2222 Camden Court
Oakbrook, Illinois 60521

Dr. Diana R. Nelson
Systems Research Labs, Inc.
2800 Indian Ripple Road
Dayton, Ohio 45440

Dr. Michael A. Nelson
Systems Research Labs, Inc.
2800 Indian Ripple Road
Dayton, Ohio 45440

Richard L. Newman
Crew Systems Consultants

Capt. Hal Nord, President
Aviation Management Advisor's Inc.
41 South Road
Rye Beach, N.H. 03871

Dr. John A. Novak
Asst. Professor, Science Education
The The Ohio State University
283 ARPS Hall
1945 N. High Street
Columbus, Ohio 43210

Herman F. W. Oberkrom, Senior Staff Engineer
Systems Group of TRW, Inc.
1340 Woodman Drive
Dayton, Ohio 45432

Dr. Karl Olson
Department of Electrical Engineering
The Ohio State University

Capt. Harry W. Orlady, (Ret.)
Orlady Associates, Inc.
312 So. Park Road
Lagrange, Il. 60525

Dr. Dean H. Owen
The Ohio State University
Department of Psychology
404-C W. 17th Ave.
Columbus, Ohio 43210

Dave Park
Aviation Psychology Lab
The Ohio State University

Dr. James Parker, President
Biotechnology, Inc.

Dr. H. McIlvaine Parsons
Human Resources Research Organization
300 N. Washington St.
Alexandria, VA 22314

Dr. Richard Pepler, V.P.
Dunlap & Associates, Inc.

David Po-Chedley
Douglas Aircraft Co.
Mail code 35-36
3855 Lakewood Blvd.
Long Beach, CA 90846

Kenneth Porada, Asst. Professor Psychology
Heidelberg College
Tiffin, Ohio 44883

Dr. Raymond E. Reilly
Allen Corporation of America
401 Wythe St.
Alexandria, VA 22314

Dr. John Reising
Flight Dynamics Laboratory
WPAFB, Ohio 45433

Dr. W. Ann Reynolds
Provost
The Ohio State University

Ira J. Rimson, P.E.
System Safety Associates, Ltd.
P.O. Box 1015
Springfield, Virginia 22151

George Rings
Department of Aviation
The Ohio State University

Ms. Linda Orlady Rings
Department of Aviation
The Ohio State University

Robert Ripley
Traveling Aviation Seminars
Dublin, OH

Dr. Malcolm Ritchie
FAA Technical Center
Atlantic City Airport, NJ

D. B. Robinson
United Airlines

Lt. Col. Joe C. Robinson
USAF
OLAA, 57 FWW,
Luke AFB, AZ 85309

Dr. Tom Rockwell
Department of Industrial and Systems Engineering
The Ohio State University

Dr. Stan Roscoe
New Mexico State University
Las Cruces, N. M. 88003

Dr. David Rotker
Venezuelan Air Force
Apartado 60096 CHACAO
Caracas 1060 Venezuela

Dr. Bill Rouse, Chairman
Mechanical and Industrial Engineering
University of Illinois
Urbana/Champaign

George E. Rowland, President
Rowland & Company
P.O. Box 61
Haddonfield, N. J. 08033

Dr. Frank T. Ruggiero
Boeing
407 Anacortes Ave., S.E.
Renton, Washington 98955

Diane Rush
Aviation Psychology Lab
The Ohio State University

Daniel W. Sauer - Research Scientist
Systems Research Laboratories, Inc.
2800 Indian Ripple Rd.
Dayton, Ohio 45440

Ms. Diane Rush
The Ohio State University

Daniel W. Sauer, Research Scientist
Systems Research Laboratories, Inc.
2800 Indian Ripple Road
Dayton, Ohio 45440

Richard J. Schiffler
ASD/ENECH Human Factors Branch
Wright Patterson AFB, Ohio

Lowell Schipper
Bowling Green State University
Department of Psychology
Bowling Green, Ohio 43403

Dr. Michael Schukert
Department of Aviation
The Ohio State University

Ron Schwartz-Crew Station Design Engineer
USAF Aeronautical Systems Div., WPAFB
ASD/ENECC
WPAFB, OH 45433

Dr. John Senders
University of Toronto
Toronto, Ontario
Canada

Ralph Sewell
Executive Air Fleet Corporation
118 Billy Diehl Road
Teterboro, New Jersey 07608

John Sheppard, ABD
Seville Research Corp.
2814 Newton Ct/
Blacksburg, VA 24060

Robin L. Shew-Research Scientist B
Systems Research Laboratories, Inc.
2800 Indian Ripple Road
Dayton, Ohio 45440

John C. Simons, Senior Scientist--Pilot, Operations
General Physics Corporation
Suite 240
1010 Woodman Drive
Dayton, Ohio 45432

Robert P. Smerke
Lockheed California Company
P.O. Box 551
Burbank, California 91520

Dave Smith
Aviation Psychology Lab
The Ohio State University

Dr. Marion Smith, Associate Dean
College of Engineering
The Ohio State University

Dr. Phil Smith
The Ohio State University
Department of Ind. & Systems Eng.
242 Baker Systems
1971 Neil Ave., Cols., Oh 43220

Dr. Walter Sperr
Consultant Mech.Engineering
Sternwartestrasse 46
A - 1180 Vienna, Austria

Dr. Edward A. Stark
Singer-Link
Binghamton, N.Y. 13902

William C. Steedman, Engineering Psychologist
Aeronautical Systems Division,
WPAFB, OH 45433

Lt. Gerald R. Stoffer, Ph.D.
Naval Training Equipment Center
Human Factors Lab (N-712)
Orlando, FL. 32813

William F. Storm
USAFSAM (School of Aerospace Medicine)
Brooks AFB, Texas

Dr. Barry Strauch
Embry Riddle Aeronautical University
Regional Airport
Daytona Beach, FL. 32014

Dr. Mervin K. Strickler, Jr.
Executive Director, Federal Union, Inc.
V.P., World Aerospace Education Organization
Suite 1218
1875 Connecticut Ave., N. W.
Washington, D. C. 20009

Dr. John L. Sullivan
Suffolk University
Beacon Hill, Boston, MA 02114

Dr. Douglas Talcott
ASD/ENECH
WPAFB, Ohio 45433

Dr. Trieve A. Tanner, Research Psychologist
NASA
Washington, D. C. 20546

Dr. H. Lee Task
AFAMRL/HEF
WPAFB, Ohio 45433

Dr. Henry Taylor, Director
Institute of Aviation
University of Illinois
Savoy, IL 61874

Richard W. Taylor, Vice President
Boeing Commercial Airplane Company
P. O. Box 3707 (Mail Stop 75-11)
Seattle, Washington 98124

Donald F. Thielke, Vice-President
Flight Engineers International ASSN.
905 Sixteenth Street
Washington, D. C. 20006

Guice Tinsley
FAA
800 Independence Avenue
Washington, D.C. 20591

Connie J. Tobias
Ohio University
RT.4
Pomeroy, Ohio 45769

Dr. Stanley Trollip
University of Illinois
Urbana/Champaign

Dr. Harold VanCott
Chief Scientist
Biotechnology, Inc.

Dirk E. Van Dam, Director
Moody Aviation
P. O. Box 429
Elizabethton, TN 37643

Chuck Ventola
Department of Aviation
The Ohio State University

Ron Ventola
Department of Aviation
The Ohio State University

Gerrit J. Walhout
Chief, Human Factors Division
National Transportation Safety Board
800 Independence Avenue TE-10
Washington, D. C. 20594

Dr. Rik Warren
Department of Psychology
The Ohio State University

Raymond L. Wasson
Human Factors Engineer
Lockheed-Georgia Company
86 South Cobb Drive
Department 72-48 Zone 450
Marietta, Georgia 30063

G. Stacy Weislogel
Department of Aviation
The Ohio State University

Marshall Whiteaker
Headquarters AFLC
Engineering Directorate
WPAFB, Ohio 45433

Leon M. Wise, Director, Aviation Psych. Lab.
Heidelberg College
Tiftin, Ohio 44883

Anchard F. Zeller, Ph.D.
Hq. Inspection and Safety Center
Life Sciences Division
Norton AFB, CA 92409

AUTHOR INDEX

W.C. Allen..................51
T.L. Amerson...............182
R.I. Anderson..............306
H. Aubauer..................63
G.J. Barbato...............226
S. Baron....................62
R.A. Benel.................182
D.B. Beringer..............334
J. Berlin..................166
F. Brecke..................145
C.A. Brictson..............362
W.R. Brown.................267
W.W. Castor................326
S.L. Chappell...............94
A.P. Ciavarelli............362
D. Damos...................279
A. Diehl....................94
J.A. Dohme.................267
E.E. Eddowes...............335
A.R. Ephrath...............216
J.M. Flach..................74
G.W. Flathers, II..........120
H.C. Foushee.................1
W.C. Giffin............120,134
R.D. Gilson.................74
E. Gora....................118
L.L. Griffin...............106
R.L. Harris, Jr............216
R. Hennessy................172
L. Hitchcock................28
C. Holmes..................166
R.J. Jagacinski.............74
G. Johannsen...............171
J.M. Koonce............255,261
R. Lancroft.................62
J.P. Loomis.................38
T.M. McCloy................261
S.J. Mangold...............205
R.W. Moss..................226
R. Muralidharan.............62
R.L. Newman................119
H.W. Orlady................247
D.H. Owen..............192,205
E.A. Palmer.................94
D.A. Po-Chedley.............61
R.F. Porter.................38
J. Reising..................28
L.O. Rings.................248
T.H. Rockwell..............120
S.N. Roscoe.............21,172
G. Rothbauer...............118
S.H. Rouse.................117

W.B. Rouse.............117,171

M.G. Sanders................12

J.E. Schofield.............134

J.W. Senders................12

T. Smist...................279

W. Sperr....................63

E.A. Stark.................289

A.T. Stephens..............216

G.R. Stoffer...............343

B. Strauch.................376

H.L. Task..................106

J.R. Tole..................216

S.R. Trollip...........306,316

R. Warren..................205

W.L. Welde.................119

A.M. Williams..............362

G. Zacharias................62